



# Regression asymptotics using martingale convergence methods.

## Citation

Ibragimov, Rustam and Peter C.B. Phillips. 2008. Regression asymptotics using martingale convergence methods. *Econometric Theory* 24(4): 888-947.

## Published Version

<http://dx.doi.org/10.1017/S0266466608080365>

## Permanent link

<http://nrs.harvard.edu/urn-3:HUL.InstRepos:2624459>

## Terms of Use

This article was downloaded from Harvard University's DASH repository, and is made available under the terms and conditions applicable to Other Posted Material, as set forth at <http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA>

## Share Your Story

The Harvard community has made this article openly available.  
Please share how this access benefits you. [Submit a story](#).

[Accessibility](#)

# REGRESSION ASYMPTOTICS USING MARTINGALE CONVERGENCE METHODS

Rustam Ibragimov

*Department of Economics, Yale University*

Peter C. B. Phillips

*Cowles Foundation for Research in Economics, Yale University*

*University of Auckland and University of York*

## ABSTRACT

Weak convergence of partial sums and multilinear forms in independent random variables and linear processes and their nonlinear analogues to stochastic integrals now plays a major role in nonstationary time series and has been central to the development of unit root econometrics. The present paper develops a new and conceptually simple method for obtaining such forms of convergence. The method relies on the fact that the econometric quantities of interest involve discrete time martingales or semimartingales and shows how in the limit these quantities become continuous martingales and semimartingales. The limit theory itself uses very general convergence results for semimartingales that were obtained in the work of Jacod and Shiryaev (2003). The theory that is developed here is applicable in a wide range of econometric models and many examples are given.

One notable outcome of the new approach is that it provides a unified treatment of the asymptotics for stationary, explosive, unit root, and local to unity autoregression, as well as some general nonlinear time series regressions. All these cases are subsumed within the martingale convergence approach and different rates of convergence are accommodated in a natural way. Moreover, the results on multivariate extensions developed in the paper deliver a unification of the asymptotics for, among many others, models with cointegration as well as for regressions with regressors that are nonlinear transforms of integrated time series driven by shocks correlated with the equation errors. Since this is the first time the methods have been used in econometrics, the exposition is presented in some detail with illustrations of new derivations of some well-known existing results, in addition to the provision of new results and the unification of the limit theory for autoregression.

---

An extended version of this work is available as Ibragimov and Phillips (2004). We are particularly grateful to three anonymous referees and Pentti Saikkonen for many helpful comments and suggestions on earlier versions. We also thank Victor Chernozhukov, Jin-Chuan Duan, Xavier Gabaix, Patrik Guggenberger, Alex Maynard, Roger Moon, Marcelo Moreira and seminar participants at Yale University, the University of California at Los Angeles, the University of Southern California, the University of Toronto and the 9th World Congress of the Econometric Society (London, 2005) for commenting on this work. The paper has also benefited from discussions with colleagues at the Departments of Economics at the University of British Columbia, the University of California at San Diego, Harvard University, the London School of Economics and Political Science, Massachusetts Institute of Technology, the Université de Montréal, McGill University and New York University, the Division of the Humanities and Social Sciences at California Institute of Technology, and Nuffield College, University of Oxford. Rustam Ibragimov gratefully acknowledges financial support from a Yale University Dissertation Fellowship and a Cowles Foundation Prize. Peter C. B. Phillips acknowledges partial research support from a Kelly Fellowship and the NSF under grant SES 04-142254. Correspondence to: peter.phillips@yale.edu

*Key words and phrases:* time series autoregression, semimartingale, martingale, convergence, stochastic integrals, bilinear forms, multilinear forms,  $U$ -statistics, unit root, stationarity, nonlinear models, Brownian motion, invariance principle, unification.

*JEL Classification:* C13, C14, C32

March, 2005

Revised version: September 22, 2006; Second Revision: March, 2007.

## 1. Introduction

Much of the modern literature on asymptotic theory in statistics and econometrics involves the weak convergence of multilinear forms and  $U$ -statistics in independent random variables (r.v.'s), martingale-differences and weakly dependent innovations to stochastic integrals (see, among others, Dynkin and Mandelbaum, 1983, Mandelbaum and Taqqu, 1984, Phillips, 1987a & b, Avram, 1988, and Borodin and I. A. Ibragimov, 1995). In econometrics, the interest in this limit theory is frequently motivated by its many applications in regression asymptotics for processes with autoregressive roots at or near unity (*inter alia*, see Phillips, 1987a & b, Phillips and Perron, 1988, Park and Phillips, 1999, 2001, Phillips and Magdalinos, 2006, and references therein). Recent attention (Park and Phillips, 1999, 2001, de Jong, 2002, Jeganathan, 2003, 2004, Pöetscher 2004, Saikkonen and Choi, 2004) has also been given to the limit behavior of certain types of nonlinear functions of integrated processes. Results of this type have interesting econometric applications that include transition behavior between regimes and market intervention policy (Hu and Phillips, 2004), where nonlinearities of nonstationary economic time series arise in a natural way.

Traditionally, functional limit theorems for multilinear forms have been derived by using their representation as polynomials in sample moments (via summation by parts arguments or, more generally, Newton polynomials relating sums of powers to the sums of products) and then applying standard weak convergence results for sums of independent or weakly dependent r.v.'s or martingales. Avram (1988), for example, makes extensive use of this approach. Thus, in the case of a martingale-difference sequence  $(\epsilon_t)$  with  $E(\epsilon_t^2|\mathfrak{F}_{t-1}) = \sigma_\epsilon^2$  for all  $t$  and  $\sup_{t \in \mathbf{Z}} E(|\epsilon_t|^p|\mathfrak{F}_{t-1}) < \infty$  a.s. for some  $p > 2$ , Donsker's theorem for the partial sum process (see Theorem 2.1), viz.,

$$\frac{1}{\sqrt{n}} \sum_{t=1}^{[nr]} \epsilon_t \rightarrow_d \sigma_\epsilon W(r),$$

where  $W = (W(s), s \geq 0)$  denotes standard Brownian motion, implies that the bilinear form

$$\frac{1}{n} \sum_{t=2}^{[nr]} \left( \sum_{i=1}^{t-1} \epsilon_i \right) \epsilon_t$$

converges to the stochastic integral  $\sigma_\epsilon^2 \int_0^r W(v) dW(v)$ . This approach has a number of advantages and has been extensively used in econometric work since Phillips (1987a).

The approach also has drawbacks. One is that the approach is problem specific in certain ways. For instance, it cannot be directly used in the case of statistics like  $\sum_{t=1}^n y_{t-1} u_t$ , where  $y_t = \alpha_n y_{t-1} + u_t$ ,  $t = 1, \dots, n$ , and  $\alpha_n \rightarrow 1$  as  $n \rightarrow \infty$ , that are central to the study of local deviations from a unit root in time series regression. Of course, there are ways of making the usual functional limit theory work (Phillips, 1987b; Chan and Wei, 1987 & 1988) and even extending it to situations where the deviations are moderately distant from unity (Phillips and Magdalinos, 2006). In addition, the method cannot be directly applied in the case of sample covariance functions of random walks and innovations, like  $V_n = n^{-1/2} \sum_{t=2}^n f\left(\frac{1}{\sqrt{n}} \sum_{i=1}^{t-1} \epsilon_i\right) \epsilon_t$ , where  $f$  is a certain nonlinear function. Such sample covariances commonly arise in econometric models where nonlinear functions are introduced to smooth transitions from one regime to another (e.g., Saikkonen and Choi, 2004). To deal with such complications, one currently has to appeal to stochastic Taylor expansions and polynomial approximations to  $V_n$ . Similar to the above, the traditional methods based on functional central limit theorems and continuous mapping arguments cannot be directly applied in the case of general one- and multisample  $U$ -statistics.

At a more fundamental level, the standard approach sheds little insight into the underlying nature of limit results such as  $n^{-1} \sum_{t=2}^{[nr]} \left( \sum_{i=1}^{t-1} \epsilon_i \right) \epsilon_t \rightarrow_d \sigma_\epsilon^2 \int_0^r W(s) dW(s)$  or  $\omega^2 \int_0^r W(s) dW(s) + r\lambda$  for some

constants  $\lambda$  and  $\omega$  in the case of weakly dependent  $\epsilon_t$ . Such results are, in fact, the natural outcome of convergence of a sequence of (semi)martingales to a continuous (semi)martingale. As such, they may be treated directly in this way using powerful methods of reducing the study of semimartingale convergence to the study of convergence of its predictable characteristics. Jacod and Shiryaev (2003, hereafter JS) pioneered developments in stochastic process limit theory along these lines (see also He, Wang and Yan, 1992, hereafter HWY), but the method has so far not been used in the theory of weak convergence to stochastic integrals, nor has it yet been used in econometrics.

The asymptotic results for semimartingales obtained by JS have great generality. However, these results appear to have had little impact so far in statistics and none that we are aware of in econometrics. In part, this may be due to the fact that the book is difficult to read, contains many complex conceptualizations, and has a highly original and demanding notational system. The methods were recently applied by Coffman, Puhalskii and Reiman (1998) to study asymptotic properties of classical polling models that arise in performance studies of computer services. In this interesting paper, Coffman, Puhalskii and Reiman showed, using the JS semimartingale convergence results, that unfinished work in a queuing system under heavy traffic tends to a Bessel type diffusion. Several applications of martingale convergence results in mathematical finance are presented in Prigent (2003). We also note that the results on convergence of martingales have previously allowed to unify the convergence of row-wise independent triangular arrays and the convergence of Markov processes (see Stroock and Varadhan, 1979, Hall and Heyde, 1980, and the review in Chapter 1 in Prigent, 2003). In addition, as discussed in, e.g., Subsection 3.3 in Prigent (2003), the martingale convergence results provide a natural framework for the analysis of the asymptotics of GARCH, stochastic volatility and related models. Several related works in probability have focused on the analysis of convergence of stochastic integrals driven by processes satisfying uniform tightness conditions or their analogues and on applications of the approach to the study of weak convergence of solutions of stochastic differential equations (see Jakubowski, Mémín and Pagès, 1989, Kurtz and Protter, 1991, Mémín and Słomiński, 1991, Mémín, 2003, and the review in Subsection 1.4 in Prigent, 2003).

The present paper develops a new approach to obtaining time series regression asymptotic results using general semimartingale convergence methods. The paper shows how results on weak convergence of semimartingales in terms of the triplets of their predictable characteristics obtained in JS may be used to develop quite general asymptotic distribution results in time series econometrics and to provide a unifying principle for studying convergence to limit processes and stochastic integrals by means of semimartingale methods. The main advantage of this treatment is its generality and range of applicability. In particular, the approach unifies the proof of weak convergence of partial sums to Brownian motion with that of the weak convergence of sample covariances to stochastic integrals of Wiener processes. Beyond this, the methods can be used to develop asymptotics for time series regression with roots near unity and to study weak convergence of nonlinear functionals of integrated processes. In all these cases, the limit theory is reduced to a special case of the weak convergence of semimartingales.

For the case of a first order autoregression with martingale-difference errors, we show that an identical construction delivers a central limit theorem in the stationary case and weak convergence to a stochastic integral in the unit root case, thereby unifying the limit theory for autoregressive estimation and realizing a long-sought-after goal in time series asymptotic theory. In fact, the approach goes further and enables a unified treatment of stationary, explosive, unit root, local to unity and nonlinear cases of time series autoregression. In all these cases, normalized versions of the estimation error are represented in martingale form as a ratio  $X_n(r)/[X_n]_r^{1/2}$ , where  $X_n(r)$  is a martingale with quadratic variation  $[X_n]_r$ , and the limit theory is delivered by martingale convergence in the form  $X_n(r)/[X_n]_r^{1/2} \rightarrow_d X(r)/[X]_r^{1/2}$ , where  $X(r)$  is the limiting martingale process.<sup>1</sup> To our knowledge, no other approach to the limit theory is able to

---

<sup>1</sup>We note that since the numerator and denominator in the self-normalized martingales in this construction are of the

accomplish this. This unification is conceptually simple and attains a goal that has eluded researchers for more than two decades.

Further, our results for the nonlinear case in Section 3 deliver a unification to the analysis of asymptotics for a wide class of models involving nonlinear transforms of integrated time series. The only condition that needs to be imposed on functions of such processes in these models, in addition to smoothness, is, essentially, that they do not grow faster than a power function. This covers most of the econometric models encountered in practice. Moreover, the general results on multivariate extensions developed in Section 4 of the paper provide a unification of the asymptotics for, among others, models with cointegration as well as for regressions with regressors that are general nonlinear transforms of integrated time series driven by shocks correlated with the equation errors.

For instance, the two asymptotic results given below in (1.1) and (1.2), which are of fundamental importance in applications, follow directly from our limit theory (see Theorems 4.1-4.3).

Suppose that  $w_t = (u_t, v_t)'$  is the linear process  $w_t = G(L)\epsilon_t = \sum_{j=0}^{\infty} G_j \epsilon_{t-j}$ , with  $G(L) = \sum_{j=0}^{\infty} G_j L^j$ ,  $\sum_{j=1}^{\infty} j |G_j| < \infty$ ,  $G(1)$  of full rank, and  $\{\epsilon_t\}_{t=0}^{\infty}$  a sequence of i.i.d. mean-zero random vectors such that  $E\epsilon_0\epsilon_0' = \Sigma_{\epsilon} > 0$  and  $\max_i E|\epsilon_{i0}|^p < \infty$  for some  $p > 4$ . Then

$$\frac{1}{n} \sum_{t=2}^{[nr]} \left( \sum_{i=1}^{t-1} u_i \right) v_t \rightarrow_d r \lambda_{uv} + \int_0^r W(v) dV(v), \quad (1.1)$$

where  $(W, V) = ((W(s), V(s)), s \geq 0)$  is bivariate Brownian motion with covariance matrix  $\Omega = G(1) \Sigma G(1)$  and  $\lambda_{uv} = \sum_{j=1}^{\infty} E u_0 v_j$ .

Further, if  $f : \mathbf{R} \rightarrow \mathbf{R}$  is a twice continuously differentiable function such that  $f'$  satisfies the growth condition  $|f'(x)| \leq K(1 + |x|^{\alpha})$  for some constants  $K > 0$  and  $\alpha > 0$  and all  $x \in \mathbf{R}$ , and if  $p \geq \max(6, 4\alpha)$ , then

$$\frac{1}{\sqrt{n}} \sum_{t=2}^{[nr]} f \left( \frac{1}{\sqrt{n}} \sum_{i=1}^{t-1} u_i \right) v_t \rightarrow_d \lambda_{uv} \int_0^r f'(W(v)) dv + \int_0^r f(W(v)) dV(v). \quad (1.2)$$

As we will show, one of the inherent advantages of the martingale approach is that it allows in a natural way for differences in rates of convergence that arise in the limit theory for autoregression. In contrast, conventional approaches require separate treatments for the stationary and nonstationary cases, as is very well-known.

In addition, the present paper contributes to the asymptotic theory of stochastic processes and time series in several other ways. First, applications of the general martingale convergence results to statistics considered in this paper overcome some technical problems that have existed heretofore in the literature. For instance, the global strong majoration condition in JS that naturally appears in the study of weak convergence to a Brownian motion is not satisfied in the case of weak convergence to stochastic integrals. This failure may explain why the martingale convergence methods of JS have not so far been applied to such problems. This paper demonstrates how this difficulty can be overcome by means of localized versions of general semimartingale results in JS that involve only a local majoration argument. These new arguments appear in the proofs of the results in Sections 3 and 4.

---

same order, the approach developed in the paper applies irrespective of particular scaling factors which may be used and these are therefore irrelevant to the limit theory and to applications of the method. Indeed, our approach will deliver the asymptotics for various discrete time martingales in the numerator of the above construction and, by means of this derivation, also deliver convergence of the denominator. The latter is, essentially, the required condition on the convergence of the second characteristic of the martingale  $X_n$ .

Second, we provide general sufficient conditions for the assumptions of JS semimartingale convergence theorems to be satisfied for multivariate diffusion processes, including the case of stochastic integrals considered in the paper (see Section 10 and, in particular, Corollary 10.1). These results provide the key to the analysis of convergence to stochastic integrals and, especially, to the study of the asymptotics of functionals of martingales and linear processes in Sections 3 and 4. Third, the general approach developed in this paper can be applied in a number of other fields of statistics and econometrics, where convergence to Gaussian processes and stochastic integrals arise. These areas include, for instance, the study of convergence of multilinear forms, nonlinear statistics and general (possibly multisample)  $U$ -statistics to multiple stochastic integrals as well as the analysis of asymptotics for empirical copula processes, all of which are experiencing growing interest in econometric research.

The paper is organized as follows. Section 2 contains applications of the approach to partial sums and sample covariances of independent r.v.'s and linear processes. Sections 3 present the paper's first group of main results, giving applications of semimartingale limit theorems to weak convergence to stochastic integrals. We obtain the asymptotic results for general classes of nonlinear functions of integrated processes and discuss their corollaries in the linear case of sample autocorrelations of linear processes and their partial sums. Section 4 provides extensions to multivariate cases, including new proofs of weak convergence to multivariate stochastic integrals. This section gives results on weak convergence of discontinuous martingales (arising from discrete time martingales) to continuous martingales and completes the unification of the limit theory for autoregression. Section 5 applies the results obtained in the paper to stationary autoregression and unit root regression. Section 6 provides an explicit unified formulation of the limit theory for first order autoregression including the case of explosive autoregression which can also be handled by martingale methods. Section 7 concludes and mentions some further applications of the new techniques.

Sections 8-12 are appendices that contain definitions and technical results needed for the arguments in the body of the paper. These appendices are intended to provide enough background material to make the body of the paper accessible to econometric readers and constitute a self-contained resource for the main stochastic process theory used here. In particular, Section 8 reviews definitions of fundamental concepts used throughout the paper. Section 9 discusses the general JS results for convergence of semimartingales in terms of their predictable characteristics. Section 10 presents sufficient conditions for semimartingale convergence theorems to hold in the case where the limit semimartingale is a diffusion or a stochastic integral. Section 11 provides results on Skorohod embedding of martingales into a Brownian motion and rates of convergence that are needed in the asymptotic arguments. Section 12 contains some auxiliary lemmas needed for the proof of the main results.

## 2 Invariance principles (IP) for partial sums, sample variances and sample covariances

In what follows, we use standard concepts and definitions of semimartingale theory (see Section 8 for further details).

Let  $\mathbf{R}_+ = [0, \infty)$  and  $\mathbf{Z} = \{\dots, -2, -1, 0, 1, 2, \dots\}$ . Throughout the paper, we assume that stochastic processes considered are defined on the Skorohod space  $(\mathbb{D}(\mathbf{R}_+^d), \mathcal{D}(\mathbf{R}_+^d))$ , if not stated otherwise (so that the time argument of the processes is nonnegative). A limit process  $X = (X(s), s \geq 0)$  appearing in the asymptotic results is the canonical process  $X(s, \alpha) = \alpha(s)$  for the element  $\alpha = (\alpha(s), s \geq 0)$  of  $\mathbb{D}(\mathbf{R}_+^d)$  (see Section VI.1 and Hypothesis IX.2.6 in JS) and  $\mathbb{F}$  is the filtration generated by  $X$ . In what follows,  $\rightarrow_d$  denotes convergence in distribution in an appropriate metric space and  $\rightarrow_P$  stands for convergence in

probability. The symbol  $=_d$  means distributional equivalence. For a sequence of r.v.'s  $\xi_n$  and constants  $a_n$ , we write  $\xi_n = O_P(1)$  if the sequence  $\xi_n$  is bounded in probability and write  $\xi_n = o_{a.s.}(a_n)$  if  $\xi_n/a_n \rightarrow_{a.s.} 0$ . As in the introduction,  $W = (W(s), s \geq 0)$  denotes standard (one-dimensional) Brownian motion on  $\mathbb{D}(\mathbf{R}_+)$ , if not stated otherwise. All processes considered in the paper are assumed to be continuous and locally square integrable, if not stated otherwise. Throughout the paper,  $K$  and  $L$  denote constants that do not depend on  $n$  (but, in general, can depend on other parameters of the settings considered) and which are not necessarily the same from one place to another.

Let  $(\epsilon_t)_{t \in \mathbf{Z}}$  be a sequence of r.v.'s and let  $(\mathfrak{F}_t)_{t \in \mathbf{Z}}$  be a natural filtration for  $(\epsilon_t)$  (that is,  $\mathfrak{F}_t$  is the  $\sigma$ -field generated by  $\{\epsilon_k, k \leq t\}$ ). The following conditions will be convenient at various points in the remainder of the paper.

**Assumption D1:**  $(\epsilon_t, \mathfrak{F}_t)$  is a martingale-difference sequence with  $E(\epsilon_t^2 | \mathfrak{F}_{t-1}) = \sigma_\epsilon^2 \in \mathbf{R}_+$  for all  $t$  and  $\sup_{t \in \mathbf{Z}} E(|\epsilon_t|^p | \mathfrak{F}_{t-1}) < \infty$  a.s. for some  $p > 2$ .

**Assumption D2:**  $(\epsilon_t)$  are mean-zero i.i.d. r.v.'s with  $E\epsilon_0^2 = \sigma_\epsilon^2 \in \mathbf{R}_+$  and  $E|\epsilon_0|^p < \infty$  for some  $p > 2$ .

The following theorems illustrate the use of the martingale convergence machinery in conjunction with the Skorohod embedding (see Appendix A4) in proving some well known martingale limit results for partial sums. In the simplest case, a sequence of discrete time martingales is embedded in a sequence of continuous martingales to which we apply martingale convergence results for continuous martingales, giving an invariance principle for martingales with non-random conditional variances. As is conventional, the proof requires that the probability space on which the random sequences are defined has been appropriately enlarged so that Lemma 11.1 in Appendix A4 holds. In the proof of the main results of the paper,  $(T_k)_{k \geq 0}$  denote the stopping times defined in Lemma 11.1.

Later in the paper in Section 4, we show how to use martingale convergence results of discontinuous martingales (semimartingales) to continuous martingales (semimartingales) which avoid the use of the Skorohod embedding. In doing so, these results are particularly useful in multivariate extensions.

**Theorem 2.1 (IP for martingales).** *Under assumption (D1),*

$$\frac{1}{\sqrt{n}} \sum_{t=1}^{[nr]} \epsilon_t \rightarrow_d \sigma_\epsilon W(r). \quad (2.1)$$

**Proof.** From Lemma 11.1 it follows that

$$\frac{1}{\sqrt{n}} \sum_{t=1}^{[nr]} \epsilon_t =_d W\left(\frac{T_{[nr]}}{n}\right). \quad (2.2)$$

By (11.3) and Lemma 12.3 in Appendix A5,

$$T_{[nr]}/n \rightarrow_P \sigma_\epsilon^2 r. \quad (2.3)$$

Therefore, from Lemma 12.2 it follows that  $W(T_{[nr]}/n) \rightarrow_d W(\sigma_\epsilon^2 r)$ . This and (2.2) imply (2.1). ■

The following theorem is the analogue of Theorem 2.1 for linear processes.



**Theorem 2.2 (IP for linear processes).** Suppose that  $(u_t)_{t \in \mathbf{N}}$  is the linear process  $u_t = C(L)\epsilon_t = \sum_{j=0}^{\infty} c_j \epsilon_{t-j}$ ,  $C(L) = \sum_{j=0}^{\infty} c_j L^j$ , where  $\sum_{j=1}^{\infty} j|c_j| < \infty$ ,  $C(1) \neq 0$ , and  $(\epsilon_t)_{t \in \mathbf{Z}}$  satisfy assumption **(D1)** with  $p \geq 4$ . Then

$$\frac{1}{\sqrt{n}} \sum_{t=1}^{[nr]} u_t \rightarrow_d \omega W(r), \quad (2.4)$$

where  $\omega^2 = \sigma_\epsilon^2 C^2(1)$ .

**Proof.** Using the Phillips-Solo (1992) device we get

$$u_t = C(1)\epsilon_t + \tilde{\epsilon}_{t-1} - \tilde{\epsilon}_t, \quad (2.5)$$

where  $\tilde{\epsilon}_t = \tilde{C}(L)\epsilon_t = \sum_{j=0}^{\infty} \tilde{c}_j \epsilon_{t-j}$ ,  $\tilde{c}_j = \sum_{i=j+1}^{\infty} c_i$  and  $\sum_{j=0}^{\infty} |\tilde{c}_j| < \infty$ . Consequently,

$$\sum_{t=1}^k u_t = C(1) \sum_{t=1}^k \epsilon_t + \tilde{\epsilon}_0 - \tilde{\epsilon}_k, \quad (2.6)$$

and, for all  $N \in \mathbf{N}$ ,

$$\sup_{0 \leq r \leq N} \left| \frac{1}{\sqrt{n}} \sum_{t=1}^{[nr]} u_t - C(1) \frac{1}{\sqrt{n}} \sum_{t=1}^{[nr]} \epsilon_t \right| \leq \frac{\tilde{\epsilon}_0}{\sqrt{n}} + \sup_{0 \leq r \leq N} \left| \frac{\tilde{\epsilon}_{[nr]}}{\sqrt{n}} \right| \leq 2 \max_{0 \leq k \leq nN} \left| \frac{\tilde{\epsilon}_k}{\sqrt{n}} \right|. \quad (2.7)$$

By Lemmas 12.4 and 12.6,

$$\max_{0 \leq k \leq nN} \left| \frac{\tilde{\epsilon}_k}{\sqrt{n}} \right| \rightarrow_P 0. \quad (2.8)$$

By Lemma 12.3, from relations (2.7) and (2.8) it follows that, for the Skorohod metric  $\rho$  on  $\mathbb{D}(\mathbf{R}_+)$ ,

$$\rho\left(\frac{1}{\sqrt{n}} \sum_{t=1}^{[nr]} u_t, C(1) \frac{1}{\sqrt{n}} \sum_{t=1}^{[nr]} \epsilon_t\right) \rightarrow_P 0.$$

By Lemma 12.1, this and Theorem 2.1 imply the desired result. ■

**Remark 2.1** Strong approximations to partial sums of independent r.v.'s, together with the Phillips-Solo (1992) device, allow one to obtain invariance principles under independence and stationarity assumptions with explicit rates of convergence. For instance, by the Hungarian construction (see Shorack and Wellner, 1986, and Csörgő and Horváth, 1993), if  $(\epsilon_t)_{t \in \mathbf{Z}}$  satisfy assumption (D2) with  $p > 4$ , then (on an appropriately enlarged probability space)  $\left| \frac{1}{\sqrt{n}} \sum_{t=1}^{[nr]} \epsilon_t - \sigma_\epsilon W(r) \right| = o_{a.s.}(n^{1/p-1/2})$ . According to Lemma 3.1 in Phillips (2006), if, in the assumptions of Theorem 2.2,  $(\epsilon_t)_{t \in \mathbf{Z}}$  satisfy assumption (D2) with  $p > 2q > 4$ , then  $\left| \frac{1}{\sqrt{n}} \sum_{t=1}^{[nr]} u_t - \omega W(r) \right| = o_{a.s.}(n^{1/q-1/2})$ .

A few results in the literature concern the functional speed of convergence (see Coquet and Mémin, 1994, and Subsection 1.4 in Prigent, 2003). Given a sequence of square-integrable martingales  $M_n$  converging to a Wiener process, these results provide a rate of convergence for solutions of stochastic differential equations driven by the  $M_n$  in terms of the rate of convergence of the quadratic variation  $[M_n, M_n]$  of the sequence. For instance, let  $X_n$  be the (unique) solution of the following stochastic differential equation:

$$X_{n,t} = X_0 + \int_0^t \sigma(X_{n,s-}) dM_{n,s},$$

where  $\sigma : \mathbf{R} \rightarrow \mathbf{R}$  is bounded above by a constant and Lipschitzian. Consider the (unique) solution of the following stochastic differential equation:

$$X_t = X_0 + \int_0^t \sigma(X_{s-}) dW(s).$$

Let, for two cadlag processes  $X = (X(s), s \geq 0)$  and  $(Y(s), s \geq 0)$  on  $\mathbf{R}_+^d$ ,  $\Pi(X, Y)$  denote the Lévy-Prokhorov distance between their distributions ( $\Pi(X, Y) = \Pi(P_X, P_Y)$ ) defined by  $\Pi(X, Y) = \inf\{\epsilon > 0 : \forall A \in \mathcal{D}(\mathbf{R}_+^d), P_X(A) \leq P_Y(A^\epsilon) + \epsilon\}$ , where  $A^\epsilon = \{x : \delta(A, x) < \epsilon\}$  and  $\delta(A, x) = \inf_{x' \in A} d(x, x')$ . Let  $a_n$  denote  $E(\sup_{t \leq T} |[M_n, M_n]_t - t|)$ . Then  $\Pi(M_n, W) \leq O(a_n^{1/9} |\ln(a_n)|^{1/2})$  and it can be deduced that  $\Pi(X_n, X) \leq O(a_n^{1/16})$ .

**Remark 2.2** The results obtained by Dedecker and Rio (2000) (see also Dedecker and Merlevède, 2002, Section 6 in Doukhan, 2003, Subsection 1.3.4 in Prigent, 2003, and Subsection 5.1 in Nze and Doukhan, 2004) provide functional central limit theorems that are not, in general, Gaussian. Let, as before,  $(\epsilon_t)_{t \in \mathbf{Z}}$  be a sequence of r.v.'s with  $E\epsilon_t = 0$ ,  $E\epsilon_t^2 < \infty$  and let  $(\mathfrak{F}_t)_{t \in \mathbf{Z}}$  be a natural filtration for  $(\epsilon_t)$ . Further, let  $Q : \mathbf{R}^{\mathbf{Z}} \rightarrow \mathbf{R}^{\mathbf{Z}}$  stand for the right shift operator, so that, for  $(x_t)_{t \in \mathbf{Z}} \in \mathbf{R}^{\mathbf{Z}}$  and  $n \in \mathbf{Z}$ , the  $n$ -th component of  $Q(x) \in \mathbf{R}^{\mathbf{Z}}$  is  $(Q(x))_n = x_{n+1}$ . Denote by  $\mathcal{J}$  the tail  $\sigma$ -algebra of  $Q$ -invariant Borel sets of  $\mathbf{R}^{\mathbf{Z}}$ . According to Dedecker and Rio (2000), the following result that provides the convergence to a mixture of Wiener process holds. Suppose that  $\sum_{t=0}^{\infty} \epsilon_0 E(\epsilon_t | \mathfrak{F}_0)$  is a convergent series in  $L^1$ . Then the sequence  $E(\epsilon_0^2 + 2\epsilon_0(\sum_{t=1}^n \epsilon_t) | \mathcal{J})$ ,  $n > 0$ , converges in  $L^1$  to some nonnegative and  $\mathcal{J}$ -measurable r.v.  $\eta$  and  $\frac{1}{\sqrt{n}} \sum_{t=1}^{[nr]} \epsilon_t \rightarrow_d \eta W(r)$ , where  $W$  is independent of  $\mathcal{J}$ . If the sequence  $(\epsilon_t)$  is ergodic then  $\eta$  is almost surely constant:  $\eta = E\epsilon_0^2 + 2\sum_{t=1}^{\infty} E\epsilon_0\epsilon_t$  (a.s.) and the standard Donsker theorem holds.

The following theorem gives a corresponding IP for sample covariances of martingale-difference sequences.

**Theorem 2.3 (IP for sample covariances of martingale-difference sequences).** Let  $(\epsilon_t)_{t \in \mathbf{Z}}$  satisfy assumption (D1) with  $p > 4$ . Then, for all  $m \geq 1$ ,

$$\frac{1}{\sqrt{n}} \sum_{t=1}^{[nr]} \epsilon_t \epsilon_{t+m} \rightarrow_d \sigma_\epsilon^2 W(r). \quad (2.9)$$

Throughout the rest of the paper, we will use the symbol  $\mathcal{I}$  to denote different quantities in the proofs and  $\eta_t$  will denote auxiliary sequences of r.v.'s arising in the arguments; these quantities and sequences are not necessarily the same from one place to another.

**Proof.** Construct the sequence of processes

$$\begin{aligned} M_n(s) = & \sum_{i=1}^{k-1} \left( W\left(\frac{T_i}{n}\right) - W\left(\frac{T_{i-1}}{n}\right) \right) \left( W\left(\frac{T_{i+m}}{n}\right) - W\left(\frac{T_{i+m-1}}{n}\right) \right) + \\ & \left( W\left(\frac{T_k}{n}\right) - W\left(\frac{T_{k-1}}{n}\right) \right) \left( W(s) - W\left(\frac{T_{k+m-1}}{n}\right) \right) \end{aligned} \quad (2.10)$$

for  $\frac{T_{k+m-1}}{n} < s \leq \frac{T_{k+m}}{n}$ ,  $k = 1, 2, \dots$ . Note that  $M_n$  is a continuous martingale with

$$\frac{1}{\sqrt{n}} \sum_{t=1}^{[nr]} \epsilon_t \epsilon_{t+m} =_d M_n\left(\frac{T_{[nr]+m}}{n}\right) \quad (2.11)$$

by Lemma 11.1. Using Theorem 9.2, we show that  $M_n \rightarrow_d \sigma_\epsilon W$ .

The first characteristics of  $M_n$  and  $\sigma_\epsilon W$  are identically zero:  $B_n(s) = B(s) = 0$ ,  $s \geq 0$ . The second characteristic of  $\sigma_\epsilon W$  is  $C(\sigma_\epsilon W)$ , where, for an element  $\alpha = (\alpha(s), s \geq 0)$ , of the Skorohod space  $\mathbb{D}(\mathbf{R}_+)$ ,  $C(s, \alpha) = [\sigma_\epsilon W, \sigma_\epsilon W](s, \alpha) = \sigma_\epsilon^2 s$ . The second characteristic of  $M_n$  is the process  $C_n = (C_n(s), s \geq 0)$ , where

$$C_n(s) = [M_n, M_n](s) = \sum_{i=1}^{k-1} \epsilon_i^2 \left( \frac{T_{i+m}}{n} - \frac{T_{i+m-1}}{n} \right) + \epsilon_k^2 \left( s - \frac{T_{k+m-1}}{n} \right)$$

for  $\frac{T_{k+m-1}}{n} < s \leq \frac{T_{k+m}}{n}$ ,  $k = 1, 2, \dots$ <sup>2</sup>

Condition (B1) of Theorem 9.2 is obviously satisfied with  $F(s) = \sigma_\epsilon^2 s$ . Condition (B2) of Theorem 9.2 is evidently satisfied by Theorem 10.2 (or by Remark 10.3). Conditions (B3) and (B4) of Theorem 9.2 and  $[sup - \beta]$  in (B5) are trivially met.

Next, we have, for  $\frac{T_{k+m-1}}{n} < s \leq \frac{T_{k+m}}{n}$ ,  $k = 1, 2, \dots$ ,

$$\begin{aligned} |C_n(s) - C(s, M_n)| &= |C_n(s) - \sigma_\epsilon^2 s| = \\ &= \left| \sum_{i=1}^{k-1} (\epsilon_i^2 - \sigma_\epsilon^2) \left( \frac{T_{i+m}}{n} - \frac{T_{i+m-1}}{n} \right) + (\epsilon_k^2 - \sigma_\epsilon^2) \left( s - \frac{T_{k+m-1}}{n} \right) \right|. \end{aligned} \quad (2.12)$$

Since, by (11.2), for  $N \in \mathbf{N}$ ,

$$\max_{k \geq 1} \{k : T_{k-1}/n < N\} \leq KNn \text{ a.s.} \quad (2.13)$$

for some constant  $K \in \mathbf{N}$ , condition  $[\gamma - \mathbf{R}_+]$  in (B5) holds if

$$\mathcal{I}_{1n} = \max_{1 \leq k \leq KNn} \left| \sum_{i=1}^{k-1} (\epsilon_i^2 - \sigma_\epsilon^2) \left( \frac{T_{i+m}}{n} - \frac{T_{i+m-1}}{n} \right) \right| \rightarrow_P 0 \quad (2.14)$$

and

$$\mathcal{I}_{2n} = \max_{1 \leq k \leq KNn} \left| (\epsilon_k^2 - \sigma_\epsilon^2) \left( \frac{T_{k+m}}{n} - \frac{T_{k+m-1}}{n} \right) \right| \rightarrow_P 0. \quad (2.15)$$

Evidently,

$$\begin{aligned} \mathcal{I}_{1n} &\leq \max_{1 \leq k \leq KNn} \frac{\sigma_\epsilon^2}{n} \left| \sum_{i=1}^{k-1} (\epsilon_i^2 - \sigma_\epsilon^2) \right| + \\ &\max_{1 \leq k \leq KNn} \left| \sum_{i=1}^{k-1} (\epsilon_i^2 - \sigma_\epsilon^2) \left( \frac{T_{i+m}}{n} - \frac{T_{i+m-1}}{n} - \frac{\sigma_\epsilon^2}{n} \right) \right| = \mathcal{I}_{1n}^{(1)} + \mathcal{I}_{1n}^{(2)}. \end{aligned}$$

---

<sup>2</sup>Note that the martingale  $M_n(s)$  can also be written as the stochastic integral  $M_n(s) = \int_0^s m_n(v) dW(v)$ , where  $m_n(s)$  are the step functions defined by  $m_n(s) = \epsilon_k$  for  $\frac{T_{k+m-1}}{n} < s \leq \frac{T_{k+m}}{n}$ ,  $k = 1, 2, \dots$ . This representation gives  $C_n(s) = \int_0^s m_n^2(v) dv$  (and, evidently, one has  $B_n(s) = 0$ ). Similar stochastic integral representations hold as well for other (semi)martingales constructed in the proofs of asymptotic results elsewhere in the paper.

By the assumptions of the theorem and Lemma 11.1,  $\eta_t^{(1)} = \epsilon_t^2 - \sigma_\epsilon^2$  and  $\eta_t^{(2)} = (\epsilon_t^2 - \sigma_\epsilon^2)(T_{t+m} - T_{t+m-1} - \sigma_\epsilon^2)$ ,  $t \geq 0$ , are martingale-difference sequences with  $E(\eta_0^{(1)})^2 = E(\epsilon_0^2 - \sigma_\epsilon^2)^2 < \infty$  and  $\sup_t E(\eta_t^{(2)})^2 \leq LE(\epsilon_0^2 - \sigma_\epsilon^2)^2 \sup_t E(\epsilon_t^4 | \mathcal{F}_{t-1}) < \infty$  for some constant  $L$  and all  $t$ . Therefore, from Lemma 12.5, we have  $|\mathcal{I}_{1n}^{(1)}| \rightarrow_P 0$  and  $|\mathcal{I}_{1n}^{(2)}| \rightarrow_P 0$  and thus (2.14) holds. By (11.2),

$$\max_{1 \leq k \leq KNn} \left| \frac{T_{k+m}}{n} - \frac{T_{k+m-1}}{n} \right| = o(n^{q-1}), \quad (2.16)$$

for any  $q > \max(1/2, 2/p) = 1/2$ . Since, under the assumptions of the theorem,

$$\max_{1 \leq k \leq KNn} n^{-2/p} |\epsilon_k^2 - \sigma_\epsilon^2| \rightarrow_P 0$$

by Lemma 12.4, using (2.16) with  $q \in (1/2, 1 - 2/p)$  (such a choice is possible since  $p > 4$ ), we get (2.15) and thus  $[\gamma - \mathbf{R}_+]$ .

Consequently, all the conditions of Theorem 9.2 are satisfied and we have that  $M_n \rightarrow_d \sigma_\epsilon W$ . This, together with (2.3) and (2.11) implies, by Lemma 12.2, that  $\frac{1}{\sqrt{n}} \sum_{t=1}^{[nr]} \epsilon_t \epsilon_{t+m} = M_n \left( \frac{T_{[nr]+m}}{n} \right) \rightarrow_d \sigma_\epsilon W(\sigma_\epsilon^2 r)$ , that is, (2.9) holds. ■

**Remark 2.3** *It is easy to see that, for each  $1 \leq l \leq m$  and all  $N > 0$ ,*

$$\sup_{0 \leq r \leq N} \left| \frac{1}{\sqrt{n}} \sum_{t=1}^{[nr]} \epsilon_t \epsilon_{t+l} - \frac{1}{\sqrt{n}} \sum_{t=1+l}^{[nr]} \epsilon_{t-l} \epsilon_t \right| = \sup_{0 \leq r \leq N} \left| \frac{1}{\sqrt{n}} \sum_{t=1}^{[nr]} \epsilon_t \epsilon_{t+l} - \frac{1}{\sqrt{n}} \sum_{t=1}^{[nr]-l} \epsilon_t \epsilon_{t+l} \right| \rightarrow_P 0,$$

$$\sup_{0 \leq r \leq N} \left| \frac{1}{\sqrt{n}} \sum_{t=1+l}^{[nr]} \epsilon_{t-l} \epsilon_t - \frac{1}{\sqrt{n}} \sum_{t=1+m}^{[nr]} \epsilon_{t-l} \epsilon_t \right| \rightarrow_P 0.$$

Using these relations, together with Lemmas 12.1 and 12.3 and the convergence results for multivariate semimartingales in Section 4 applied to the martingale  $(\frac{1}{\sqrt{n}} \sum_{t=1+m}^{[nr]} (\epsilon_t^2 - \sigma_\epsilon^2), \frac{1}{\sqrt{n}} \sum_{t=1+m}^{[nr]} \epsilon_{t-1} \epsilon_t, \dots, \frac{1}{\sqrt{n}} \sum_{t=1+m}^{[nr]} \epsilon_{t-m} \epsilon_t)$ , one can skip the Skorohod embedding argument in the proof of Theorem 2.3. It is also not difficult to show, similar to the arguments in Theorems 4.1 and 4.2, that the following joint convergence of sample variances and sample covariances holds under assumption (D2) with  $p > 4$ :

$$\left( \frac{1}{\sqrt{n}} \sum_{t=1}^{[nr]} (\epsilon_t^2 - \sigma_\epsilon^2), \frac{1}{\sqrt{n}} \sum_{t=1}^{[nr]} \epsilon_t \epsilon_{t+1}, \dots, \frac{1}{\sqrt{n}} \sum_{t=1}^{[nr]} \epsilon_t \epsilon_{t+m} \right) \rightarrow_d \left( [E(\epsilon_t^2 - \sigma_\epsilon^2)^2]^{1/2} W^0(r), \sigma_\epsilon^2 W^1(r), \dots, \sigma_\epsilon^2 W^m(r) \right)$$

for all  $m \geq 1$ , where  $(W^0(r), W^1(r), \dots, W^m(r))$  is a standard  $(m+1)$ -dimensional Brownian motion.

As is well-known (see, e.g., Phillips and Solo, 1992, Remarks 3.9), an analogue of Theorem 2.3 for sample covariances of linear processes has the form provided by the following theorem.

**Theorem 2.4 (IP for sample covariances of linear processes).** *Suppose that  $u_t$  is the linear process  $u_t = C(L)\epsilon_t = \sum_{j=0}^{\infty} c_j \epsilon_{t-j}$ ,  $C(L) = \sum_{j=0}^{\infty} c_j L^j$ , where  $\sum_{j=1}^{\infty} j c_j^2 < \infty$ ,  $C(1) \neq 0$ , and  $(\epsilon_t)_{t \in \mathbf{Z}}$  satisfy assumption (D2) with  $p > 4$ . Then, for all  $m \geq 1$ ,*

$$\frac{1}{\sqrt{n}} \sum_{t=1}^{[nr]} (u_t u_{t+m} - \gamma_m) \rightarrow_d v(m) W(r), \quad (2.17)$$

where<sup>3</sup>  $\gamma_m = g_m(1)\sigma_\epsilon^2$ ,  $v(m) = \left(g_m^2(1)E(\epsilon_0^2 - \sigma_\epsilon^2)^2 + \sum_{s=1}^\infty (g_{m+s}(1) + g_{m-s}(1))^2 \sigma_\epsilon^4\right)^{1/2}$ ,  $g_j(1) = \sum_{k=0}^\infty c_k c_{k+j}$ ,  $j \in \mathbf{Z}$ , and it is assumed that  $c_j = 0$  for  $j < 0$ .

**Proof.** Treating  $c_j$  as zero for  $j < 0$ , define the lag polynomials  $g_j(L)$ ,  $j \in \mathbf{Z}$ , by  $g_j(L) = \sum_{k=0}^\infty c_k c_{k+j} L^k = \sum_{k=0}^\infty g_{jk} L^k$ . Further, let  $\tilde{g}_j(L) = \sum_{k=0}^\infty \tilde{g}_{jk} L^k$ , where  $\tilde{g}_{jk} = \sum_{s=k+1}^\infty g_{js} = \sum_{s=k+1}^\infty c_s c_{s+j}$ . As in Remark 3.9 of Phillips and Solo (1992), we have

$$\begin{aligned} \frac{1}{\sqrt{n}} \sum_{t=1}^{[nr]} (u_t u_{t+m} - \gamma_m) &= \frac{1}{\sqrt{n}} g_m(1) \sum_{t=1}^{[nr]} (\epsilon_t^2 - \sigma_\epsilon^2) + \\ &\frac{1}{\sqrt{n}} \sum_{t=1}^{[nr]} \sum_{s=1}^\infty g_{m+s}(1) \epsilon_{t-s} \epsilon_t + \frac{1}{\sqrt{n}} \sum_{t=1}^{[nr]} \sum_{s=1}^m g_{m-s}(1) \epsilon_t \epsilon_{t+s} + \\ &\frac{1}{\sqrt{n}} \sum_{t=1}^{[nr]} \sum_{s=m+1}^\infty g_{s-m}(1) \epsilon_{t+m-s} \epsilon_{t+m} - \\ &\frac{1}{\sqrt{n}} (\tilde{u}_{a0} - \tilde{u}_{a,[nr]}) - \frac{1}{\sqrt{n}} (\tilde{u}_{b0} - \tilde{u}_{b,[nr]}), \end{aligned} \quad (2.18)$$

where

$$\tilde{u}_{at} = \tilde{g}_m(L) \epsilon_t^2$$

and

$$\tilde{u}_{bt} = \sum_{s=1}^\infty \tilde{g}_{m+s}(L) \epsilon_{t-s} \epsilon_t + \sum_{s=1}^m \tilde{g}_{m-s}(L) \epsilon_t \epsilon_{t+s} + \sum_{s=m+1}^\infty \tilde{g}_{s-m}(L) \epsilon_{t+m-s} \epsilon_{t+m}$$

(the validity of decomposition (2.18) follows from Lemma 3.6 in Phillips and Solo, 1992).

Using Remark 2.3, it is not difficult to show that

$$\begin{aligned} &\frac{1}{\sqrt{n}} g_m(1) \sum_{t=1}^{[nr]} (\epsilon_t^2 - \sigma_\epsilon^2) + \frac{1}{\sqrt{n}} \sum_{t=1}^{[nr]} \sum_{s=1}^\infty g_{m+s}(1) \epsilon_{t-s} \epsilon_t \\ &+ \frac{1}{\sqrt{n}} \sum_{t=1}^{[nr]} \sum_{s=1}^m g_{m-s}(1) \epsilon_t \epsilon_{t+s} + \frac{1}{\sqrt{n}} \sum_{t=1}^{[nr]} \sum_{s=m+1}^\infty g_{s-m}(1) \epsilon_{t+m-s} \epsilon_{t+m} \\ &\rightarrow_d v(m) W(r). \end{aligned}$$

---

<sup>3</sup>  $g_j(1)$  are the values of the lag polynomials defined in the proof.

By (2.18) and Lemmas 12.1 and 12.3, it remains to prove that, for all  $N > 0$ ,

$$\sup_{0 \leq r \leq N} \left| \frac{1}{\sqrt{n}}(\tilde{u}_{a0} - \tilde{u}_{a,[nr]}) + \frac{1}{\sqrt{n}}(\tilde{u}_{b0} - \tilde{u}_{b,[nr]}) \right| \rightarrow_P 0. \quad (2.20)$$

But this holds since, by Lemma 12.8,  $Eu_{a0}^2 < \infty$  and  $Eu_{b0}^2 < \infty$ , and, thus, according to Lemma 12.4,

$$\max_{0 \leq k \leq nN} n^{-1/2} |\tilde{u}_{a,k}| \rightarrow_P 0$$

and  $\max_{0 \leq k \leq nN} n^{-1/2} |\tilde{u}_{b,k}| \rightarrow_P 0$ . ■

### 3 Convergence to stochastic integrals

The martingale convergence approach developed in the paper can be used to derive asymptotic results for various general functionals of partial sums of linear processes. These results are particularly useful in practice for models where nonlinear functions of integrated processes arise.

**Theorem 3.1** *Let  $f : \mathbf{R} \rightarrow \mathbf{R}$  be a twice continuously differentiable function such that  $f'$  satisfies the growth condition<sup>4</sup>  $|f'(x)| \leq K(1 + |x|^\alpha)$  for some constants  $K > 0$  and  $\alpha > 0$  and all  $x \in \mathbf{R}$ . Suppose that  $u_t$  is the linear process  $u_t = C(L)\epsilon_t = \sum_{j=0}^{\infty} c_j \epsilon_{t-j}$ ,  $C(L) = \sum_{j=0}^{\infty} c_j L^j$ , where  $\sum_{j=1}^{\infty} j|c_j| < \infty$ ,  $C(1) \neq 0$ , and  $(\epsilon_t)_{t \in \mathbf{Z}}$  satisfy assumption (D2) with  $p \geq \max(6, 4\alpha)$ . Then*

$$\begin{aligned} & \frac{1}{\sqrt{n}} \sum_{t=2}^{[nr]} f \left( \frac{1}{\sqrt{n}} \sum_{i=1}^{t-1} u_i \right) u_t \rightarrow_d \\ & \lambda \int_0^r f'(\omega W(v)) dv + \omega \int_0^r f(\omega W(v)) dW(v), \end{aligned} \quad (3.1)$$

where  $\lambda = \sum_{j=1}^{\infty} Eu_0 u_j$  and  $\omega^2 = \sigma_\epsilon^2 C^2(1)$ .

Theorem 3.1 with  $f(x) = x$  implies the following corollary that provides the conventional weak convergence limit theory for the sample covariances of linear processes  $u_t$  and their partial sums to a stochastic integral that arises in a unit root autoregression. While other proofs of this result are available (using partial summation, for example), the derivation in Theorem 3.1 shows that the result may be obtained directly by a semimartingale convergence argument.

**Corollary 3.1** *Suppose that  $u_t$  is the linear process  $u_t = C(L)\epsilon_t = \sum_{j=0}^{\infty} c_j \epsilon_{t-j}$ ,  $C(L) = \sum_{j=0}^{\infty} c_j L^j$ , where  $\sum_{j=1}^{\infty} j|c_j| < \infty$ ,  $C(1) \neq 0$ , and  $(\epsilon_t)_{t \in \mathbf{Z}}$  satisfy assumption (D2) with  $p > 4$ . Then*

$$\frac{1}{n} \sum_{t=2}^{[nr]} \left( \sum_{i=1}^{t-1} u_i \right) u_t \rightarrow_d r\lambda + \omega^2 \int_0^r W(v) dW(v), \quad (3.2)$$

where  $\lambda = \sum_{j=1}^{\infty} Eu_0 u_j$  and  $\omega^2 = \sigma_\epsilon^2 C^2(1)$ .

---

<sup>4</sup>This assumption evidently implies that  $f$  satisfies a similar growth condition with the power  $1 + \alpha$ , i.e.,  $|f(x)| \leq K(1 + |x|^{1+\alpha})$  for some constant  $K$  and all  $x \in \mathbf{R}$ .

**Remark 3.1** The processes on the right-hand side of (3.1) belong to an important class of limit semimartingales for functionals of partial sums of linear processes whose first predictable characteristics (the drift terms) are non-deterministic. The latter is a qualitative difference between the semimartingales in (3.1) and the processes on the right-hand side of (3.2), where the first characteristics are deterministic ( $r\lambda, r \geq 0$ ).

**Remark 3.2** From the proof of Theorem 3.1 it follows that the assumption that  $f$  is twice continuously differentiable can be replaced by the condition that  $f$  has a locally Lipschitz continuous first derivative, that is, for every  $N \in \mathbf{N}$  there exists a constant  $K_N$  such that  $|f'(x) - f'(y)| \leq K_N|x - y|$  for all  $x, y \in \mathbf{R}$  with  $|x| \leq N$  and  $|y| \leq N$ .

**Remark 3.3** From the proof of Theorem 3.1 we find that the following extension holds. Let  $f : \mathbf{R} \rightarrow \mathbf{R}$  be a twice continuously differentiable function such that  $f'$  satisfies the growth condition  $|f'(x)| \leq K(1 + |x|^\alpha)$  for some constants  $K > 0$  and  $\alpha > 0$  and all  $x \in \mathbf{R}$ . Suppose that  $u_t$  and  $v_t$  are two linear processes:  $u_t = \Gamma(L)\epsilon_t = \sum_{j=0}^{\infty} \gamma_j \epsilon_{t-j}$ ,  $v_t = \Delta(L)\epsilon_t = \sum_{j=0}^{\infty} \delta_j \epsilon_{t-j}$ ,  $\Gamma(L) = \sum_{j=0}^{\infty} \gamma_j L^j$ ,  $\Delta(L) = \sum_{j=0}^{\infty} \delta_j L^j$ , where  $\sum_{j=1}^{\infty} j|\gamma_j| < \infty$ ,  $\sum_{j=1}^{\infty} j|\delta_j| < \infty$ ,  $\Gamma(1) \neq 0$ ,  $\Delta(1) \neq 0$ , and  $(\epsilon_t)_{t \in \mathbf{Z}}$  satisfy assumption (D2) with  $p \geq \max(6, 4\alpha)$ . Then,

$$\frac{1}{\sqrt{n}} \sum_{t=2}^{[nr]} f \left( \frac{1}{\sqrt{n}} \sum_{i=1}^{t-1} u_i \right) u_t \rightarrow_d \lambda_{uv} \int_0^r f'(\omega_u W(v)) dv + \omega_v \int_0^r f(\omega_u W(v)) dW(v),$$

where  $\omega_u^2 = \sigma_\epsilon^2 \Gamma^2(1)$ ,  $\omega_v^2 = \sigma_\epsilon^2 \Delta^2(1)$  and  $\lambda_{uv} = \sum_{j=1}^{\infty} E u_0 v_j$ .

In particular, in the unit root case with  $f(x) = x$  we get that if  $(\epsilon_t)_{t \in \mathbf{Z}}$  satisfy assumption (D2) with  $p > 4$ , then

$$\frac{1}{n} \sum_{t=2}^{[nr]} \left( \sum_{i=1}^{t-1} u_i \right) v_t \rightarrow_d r \lambda_{uv} + \omega_u \omega_v \int_0^r W(v) dW(v),$$

where  $\omega_u^2 = \sigma_\epsilon^2 \Gamma^2(1)$ ,  $\omega_v^2 = \sigma_\epsilon^2 \Delta^2(1)$  and  $\lambda_{uv} = \sum_{j=1}^{\infty} E u_0 v_j$ .

One should also note that, as follows from the proof of Theorem 3.1, if  $\epsilon_t$  satisfies assumption (D1) with  $p > 6$  (so that  $\lambda = \sum_{j=1}^{\infty} E \epsilon_0 \epsilon_j = 0$ ), then the relation

$$\frac{1}{\sqrt{n}} \sum_{t=2}^{[nr]} f \left( \frac{1}{\sqrt{n}} \sum_{i=1}^{t-1} \epsilon_i \right) \epsilon_t \rightarrow_d \sigma_\epsilon \int_0^r f(\sigma_\epsilon W(v)) dW(v)$$

holds if  $f$  satisfies the exponential growth condition  $|f(x)| \leq 1 + \exp(K|x|)$  for some constant  $K > 0$  and all  $x \in \mathbf{R}$ . One can also deduce from the proof that the convergence

$$\frac{1}{n} \sum_{t=2}^{[nr]} \left( \sum_{i=1}^{t-1} \epsilon_i \right) \epsilon_t \rightarrow_d \sigma_\epsilon^2 \int_0^r W(v) dW(v). \quad (3.3)$$

in the case  $f(x) = x$  holds if  $\epsilon_t$  satisfies assumption (D1) with  $p > 4$ .

**Remark 3.4** Some existing results available in the literature (see Jakubowski, Mémmin and Pages, 1989, Kurtz and Protter, 1991, and Subsection 1.4 in Prigent, 2003) on convergence to stochastic integrals can

be applied to obtain convergence results such as (3.3). For instance, denote

$$N_{n,r} = \frac{\epsilon_0}{\sqrt{n}} + \sum_{t=1}^{\lfloor nr \rfloor} \frac{\epsilon_t}{\sqrt{n}}.$$

Assumption D1 implies that  $N_{n,r}$  is square integrable martingale. Since, clearly, the following stochastic integral representation holds for the statistic on the right-hand side of (3.3):

$$\frac{1}{n} \sum_{t=2}^{\lfloor nr \rfloor} \left( \sum_{i=1}^{t-1} \epsilon_i \right) \epsilon_t = \int_0^r N_{n,s} dN_{n,s},$$

asymptotic relation (3.3) can be deduced from, e.g., Theorem 2.6 of Jakubowski, Mémmin and Pages (1989) (see also Theorem 1.4.3 in Prigent, 2003). Since  $N_{n,r} \rightarrow_d W(r)$ , the latter result implies that (3.3) holds provided that the sequence of processes  $\{N_{n,r}\}_n$  satisfies the uniform tightness condition. On the other hand, from Proposition 3.2 (a) in Jakubowski, Mémmin and Pagès (1989) (part (2) of Theorem 1.4.2 in Prigent, 2003) it follows that the uniform tightness condition for  $\{N_{n,r}\}_n$  holds provided that  $\sup_n E(\sup_{r \leq t} |\Delta N_{n,r}|) < \infty$  for all  $t < \infty$ . We have that

$$E(\sup_{r \leq t} |\Delta N_{n,r}|) \leq 2E(\sup_{r \leq t} |N_{n,r}|) = 2E\left(\max_{0 \leq k \leq \lfloor nt \rfloor} \left| \frac{\epsilon_0}{\sqrt{n}} + \sum_{t=1}^k \frac{\epsilon_t}{\sqrt{n}} \right|\right).$$

By the Burkholder inequality for martingales (see Burkholder, 1973, Hall and Heyde, 1980, and de la Peña, Ibragimov and Sharakhmetov, 2003),

$$\begin{aligned} E\left(\max_{0 \leq k \leq \lfloor nt \rfloor} \left| \sum_{i=0}^k \frac{\epsilon_i}{\sqrt{n}} \right|^p\right) &\leq K_p \left\{ E\left(\frac{1}{n} \sum_{i=0}^{\lfloor nt \rfloor} E(\epsilon_i^2 | \mathfrak{F}_{i-1})\right)^{p/2} + \frac{1}{n^{p/2}} \sum_{i=0}^{\lfloor nt \rfloor} E|\epsilon_i|^p \right\}, \\ E\left(\max_{0 \leq k \leq \lfloor nt \rfloor} \left| \sum_{i=0}^k \frac{\epsilon_i}{\sqrt{n}} \right|^p\right) &\leq K_p \left\{ E\left(\frac{1}{n} \sum_{i=0}^{\lfloor nt \rfloor} E(\epsilon_i^2 | \mathfrak{F}_{i-1})\right)^{p/2} + \frac{1}{n^{p/2-1}} \max_{i \leq \lfloor nt \rfloor} E|\epsilon_i|^p \right\}, \end{aligned} \quad (3.4)$$

where  $K_p$  is a constant depending only on  $p$ . This, together with Jensen's inequality implies that, under assumption (D1), the right-hand side of (3.4) is bounded by a constant that does not depend on  $n$  and, thus,  $\sup_n E(\sup_{r \leq t} |\Delta N_{n,r}|) < \infty$  for all  $t < \infty$ . According to the above discussion, this implies that (3.3) indeed holds.

**Remark 3.5** The assumption  $|f'(x)| \leq K(1 + |x|^\alpha)$ , together with the moment condition  $E|\epsilon_0|^p < \infty$  for  $p > \max(6, 4\alpha)$ , guarantees, by Lemma 12.12, that bound (12.12) for moments of partial sums in Appendix A5 holds. As follows from the proof, Theorem 3.1 in fact holds for  $p \geq 6$  and all twice continuously differentiable functions  $f$  for which the estimate (12.12) is true and  $f'$  (and, thus,  $f$  itself) satisfies the exponential growth condition  $|f'(x)| \leq 1 + \exp(K|x|)$  for some constant  $K > 0$  and all  $x \in \mathbf{R}$ .

**Remark 3.6** Let  $X_t$  be a (nonstationary) fractional process generated by the model  $(1 - L)^d X_t = u_t$ ,  $d > 1/2$ ,  $t = 0, 1, 2, \dots$ , where  $u_t = C(L)\epsilon_t = \sum_{j=0}^{\infty} c_j \epsilon_{t-j}$  for  $t \geq 1$ ,  $u_t = 0$  for  $t \leq 0$ ,  $C(L) = \sum_{j=0}^{\infty} c_j L^j$ ,  $\sum_{j=1}^{\infty} j|c_j| < \infty$ ,  $C(1) \neq 0$ , and  $(\epsilon_t)_{t \in \mathbf{Z}}$  satisfy assumption (D2) with  $p > \max(\frac{2}{2d-1}, 2)$  (see Phillips, 1999, and Doukhan, Oppenheim and Taqqu, 2003). There are analogues of Theorems 3.1 and 3.1 for suitably normalized statistics of the long memory time series  $X_t$ . The argument is much simpler in the present



instance since the analogues of the theorems are consequences of the continuous mapping theorem and the following IP for  $X_t$  given by Lemma 3.4 in Phillips (1999a) (see Aksom and Gouriéroux, 1987, for the case of stationary ARMA components  $u_t$ ):

$$\frac{X_{[nr]}}{n^{d-1/2}} \rightarrow_d \omega^2 W_{d-1}(r) = \frac{\omega^2}{\Gamma(d)} \int_0^r (r-s)^{d-1} dW(s), \quad (3.5)$$

where  $\omega^2 = \sigma_\epsilon^2 C^2(1)$  and  $\Gamma(d) = \int_0^\infty x^{d-1} e^{-x} dx$ . Using the continuous mapping theorem, we conclude from (3.5) that the following analogues of relations (2.4), (3.2) and (3.1) hold for partial sums of elements of the fractionally integrated process  $X_t$ :

$$\frac{1}{n^{d+1/2}} \sum_{t=1}^{[nr]} X_t \rightarrow_d \omega^2 \int_0^1 W_{d-1}(r) dr,$$

$$\frac{1}{n^{2d+1}} \sum_{t=2}^{[nr]} \left( \sum_{i=1}^{t-1} X_i \right) X_t \rightarrow_d \omega^4 \int_0^r \left( \int_0^s W_{d-1}(t) dt \right) W_{d-1}(s) ds,$$

$$\frac{1}{n^{d+1/2}} \sum_{t=2}^{[nr]} f \left( \frac{1}{n^{d+1/2}} \sum_{i=1}^{t-1} X_i \right) X_t \rightarrow_d \omega^2 \int_0^r f \left( \omega^2 \int_0^s W_{d-1}(t) dt \right) W_{d-1}(s) ds,$$

where  $f$  is a continuous function. Similar functional limit theorems for discrete Fourier transforms of fractional processes can be obtained (see Phillips, 1999).

**Proof.** We first show that

$$\begin{aligned} \mathcal{I}_n &= \frac{\lambda}{n} \sum_{t=2}^{[nr]} f' \left( \frac{1}{\sqrt{n}} \sum_{i=1}^{t-1} u_i \right) + \frac{C(1)}{\sqrt{n}} \sum_{t=2}^{[nr]} f \left( \frac{1}{\sqrt{n}} \sum_{i=1}^{t-1} u_i \right) \epsilon_t \rightarrow_d \\ &\quad \lambda \int_0^r f'(\omega W(v)) dv + \omega \int_0^r f(\omega W(v)) dW(v). \end{aligned} \quad (3.6)$$

Consider the continuous semimartingale  $M_n = (M_n(s), s \geq 0)$ , where

$$\begin{aligned} M_n(s) &= \frac{\lambda}{n} \sum_{i=2}^{k-1} f' \left( \frac{1}{\sqrt{n}} \sum_{j=1}^{i-1} u_j \right) + \lambda f' \left( \frac{1}{\sqrt{n}} \sum_{j=1}^{k-1} u_j \right) \left( s - \frac{k-1}{n} \right) \\ &\quad + \sum_{i=1}^{k-1} f \left( \frac{1}{\sqrt{n}} \sum_{j=1}^{i-1} u_j \right) \left( W \left( \frac{T_i}{n} \right) - W \left( \frac{T_{i-1}}{n} \right) \right) \\ &\quad + f \left( \frac{1}{\sqrt{n}} \sum_{j=1}^{k-1} u_j \right) \left( W(s) - W \left( \frac{T_{k-1}}{n} \right) \right), \end{aligned} \quad (3.7)$$

for  $\frac{T_{k-1}}{n} < s \leq \frac{T_k}{n}$ ,  $k = 1, 2, \dots$ . By Lemma 11.1, we have the following semimartingale representation for the left-hand side of (3.6) :

$$\mathcal{I}_n =_d M_n \left( \frac{T_{[nr]}}{n} \right). \quad (3.8)$$

Further, let  $X_n = (X_n(s), s \geq 0)$  for  $n \geq 1$  and  $X = (X(s), s \geq 0)$  be the continuous vector martingales with

$$X_n(s) = (M_n(s), W(s))$$

and

$$X(s) = (h_0(1) \int_0^s f'(C(1)W(v))dv + \int_0^s f(C(1)W(v))dW(v), W(s)),$$

where

$$\lambda = h_0(1)\sigma_\epsilon^2. \quad (3.9)$$

The first characteristic of  $X_n$  is the process  $(B_n(s), s \geq 0)$ , where

$$B_n(s) = \left( \frac{\lambda}{n} \sum_{i=2}^{k-1} f' \left( \frac{1}{\sqrt{n}} \sum_{j=1}^{i-1} u_j \right) + \lambda f' \left( \frac{1}{\sqrt{n}} \sum_{j=1}^{k-1} u_j \right) \left( s - \frac{k-1}{n} \right), 0 \right) =$$

$$(B_n^1(s), B_n^2(s)) \quad (3.10)$$

for  $\frac{T_{k-1}}{n} < s \leq \frac{T_k}{n}$ ,  $k = 1, 2, \dots$ . The second characteristic of  $X_n$  is the process  $C_n = (C_n(s), s \geq 0)$  with

$$C_n(s) = \begin{pmatrix} C_n^{11}(s) & C_n^{12}(s) \\ C_n^{21}(s) & C_n^{22}(s) \end{pmatrix}, \quad (3.11)$$

where

$$C_n^{11}(s) = \sum_{i=2}^{k-1} f^2 \left( \frac{1}{\sqrt{n}} \sum_{j=1}^{i-1} u_j \right) \left( \frac{T_i}{n} - \frac{T_{i-1}}{n} \right) +$$

$$f^2 \left( \frac{1}{\sqrt{n}} \sum_{j=1}^{k-1} u_j \right) \left( s - \frac{T_{k-1}}{n} \right), \quad (3.12)$$

$$C_n^{12}(s) = C_n^{21}(s) = \sum_{i=2}^{k-1} f \left( \frac{1}{\sqrt{n}} \sum_{j=1}^{i-1} u_j \right) \left( \frac{T_i}{n} - \frac{T_{i-1}}{n} \right) +$$

$$f \left( \frac{1}{\sqrt{n}} \sum_{j=1}^{k-1} u_j \right) \left( s - \frac{T_{k-1}}{n} \right), \quad (3.13)$$

for  $\frac{T_{k-1}}{n} < s \leq \frac{T_k}{n}$ ,  $k = 1, 2, \dots$ , and

$$C_n^{22}(s) = s. \quad (3.14)$$

The process  $X$  is a solution to stochastic differential equation (10.6) with  $g_1(x) = f(C(1)x)$ ,  $x \in \mathbf{R}$ , and  $g_2(x) = h_0(1)f'(C(1)x)$ ,  $x \in \mathbf{R}$ . The first and second predictable characteristics of  $X$  are, respectively,  $B(X)$  and  $C(X)$ , where  $B$  and  $C$  are defined in (10.7) with the above  $g_i(x)$ ,  $i = 1, 2$ .

As in the proof of Theorems 3.3 and 3.1, we proceed to show that  $X_n \rightarrow_d X$  by verifying the conditions of Theorem 9.1 in order.

For  $x \in \mathbf{R}$ , let  $x_+ = \max(x, 0)$  and  $x_- = \max(-x, 0)$  and let  $B^i(s, \alpha)$ ,  $i = 1, 2$ , and  $C^{ij}(s, \alpha)$ ,  $1 \leq i, j \leq 2$ , be as in (10.7) with  $g_1(x) = f(C(1)x)$  and  $g_2(x) = h_0(1)f'(C(1)x)$ . Since, obviously,  $B^1(s, \alpha) = \int_0^s [h_0(1)f'(C(1)\alpha_2(v))]_+ dv - \int_0^s [h_0(1)f'(C(1)\alpha_2(v))]_- dv$  for  $\alpha = ((\alpha_1(s), \alpha_2(s)), s \geq 0) \in \mathbb{D}(\mathbf{R}_+^2)$ , one has (see Definition 8.3)

$$\begin{aligned} \text{Var}(B^1)(s, \alpha) + \text{Var}(B^2)(s, \alpha) &= \int_0^s [h_0(1)f'(C(1)\alpha_2(v))]_+ dv + \\ &\int_0^s [h_0(1)f'(C(1)\alpha_2(v))]_- dv = \int_0^s |h_0(1)f'(C(1)\alpha_2(v))| dv = H(s, \alpha). \end{aligned}$$

Let  $0 \leq r < s$ . For the stopping time  $S^a(\alpha)$  defined in (9.1) and for all  $v \in (r \wedge S^a(\alpha), s \wedge S^a(\alpha))$  we have  $|\alpha_2(v)| \leq |\alpha(v)| < a$  and thus  $|f(C(1)\alpha_2(v))| \leq \max_{|x| < a} |f(C(1)x)| = G_1(a)$  and  $|f'(C(1)\alpha_2(v))| \leq \max_{|x| < a} |f'(C(1)x)| = G_2(a)$ . Consequently,

$$\begin{aligned} H(s \wedge S^a(\alpha), \alpha) - H(r \wedge S^a(\alpha), \alpha) &= \int_{r \wedge S^a(\alpha)}^{s \wedge S^a(\alpha)} |h_0(1)f'(C(1)W(v))| dv \leq \\ &|h_0(1)|G_2(a)(s - r), \end{aligned} \tag{3.15}$$

$$\begin{aligned} C^{11}(s \wedge S^a(\alpha), \alpha) - C^{11}(r \wedge S^a(\alpha), \alpha) &= \\ \int_{r \wedge S^a(\alpha)}^{s \wedge S^a(\alpha)} f^2(C(1)\alpha_2(v)) dv &\leq G_1^2(a)(s - r), \end{aligned} \tag{3.16}$$

$$\begin{aligned} C^{22}(s \wedge S^a(\alpha), \alpha) - C^{22}(r \wedge S^a(\alpha), \alpha) &= \\ s \wedge S^a(\alpha) - r \wedge S^a(\alpha) &\leq (s - r). \end{aligned} \tag{3.17}$$

By (3.15)-(3.17), condition (A1) of Theorem 9.1 is satisfied with

$$F(s, a) = \max(G_1^2(a), |h_0(1)|G_2(a), 1)s.$$

Since, under assumptions of the theorem, the functions  $g_1(x) = f(C(1)x)$  and  $g_2(x) = h_0(1)f'(C(1)x)$  are locally Lipschitz continuous and satisfy growth condition (10.8), from Corollaries 10.1 and 10.2 it follows that conditions (A2)-(A4) of Theorem 9.1 hold. Condition (A5) of Theorem 9.1 is trivially satisfied since  $X_n(0) = X(0) = 0$ .

Let

$$\tilde{B}_n^1(s) = h_0(1) \sum_{i=2}^{k-1} f' \left( \frac{1}{\sqrt{n}} \sum_{j=1}^{i-1} u_j \right) \left( \frac{T_i}{n} - \frac{T_{i-1}}{n} \right) +$$

$$h_0(1)f'\left(\frac{1}{\sqrt{n}}\sum_{j=1}^{k-1}u_j\right)\left(s-\frac{T_{k-1}}{n}\right) \quad (3.18)$$

for  $\frac{T_{k-1}}{n} < s \leq \frac{T_k}{n}$ ,  $k = 1, 2, \dots$ . It is not difficult to see that

$$\sup_{0 < s \leq N} |B_n^1(s) - \tilde{B}_n^1(s)| \rightarrow_P 0. \quad (3.19)$$

Indeed, by (3.9), we have that, for  $\frac{T_{k-1}}{n} < s \leq \frac{T_k}{n}$ ,  $k = 1, 2, \dots$ ,

$$\begin{aligned} |B_n^1(s) - \tilde{B}_n^1(s)| &= \left| h_0(1) \sum_{i=2}^{k-1} f'\left(\frac{1}{\sqrt{n}} \sum_{j=1}^{i-1} u_j\right) \left(\frac{T_i}{n} - \frac{T_{i-1}}{n} - \frac{\sigma_\epsilon^2}{n}\right) \right. \\ &\quad \left. + h_0(1) f'\left(\frac{1}{\sqrt{n}} \sum_{j=1}^{k-1} u_j\right) \left(\frac{k-1}{n} \sigma_\epsilon^2 - \frac{T_{k-1}}{n}\right) \right| \\ &\leq |h_0(1)| \left| \sum_{i=2}^{k-1} f'\left(\frac{1}{\sqrt{n}} \sum_{j=1}^{i-1} u_j\right) \left(\frac{T_i}{n} - \frac{T_{i-1}}{n} - \frac{\sigma_\epsilon^2}{n}\right) \right| \\ &\quad + |h_0(1)| \left| f'\left(\frac{1}{\sqrt{n}} \sum_{j=1}^{k-1} u_j\right) \right| \left| \frac{T_{k-1}}{n} - \frac{k-1}{n} \sigma_\epsilon^2 \right|. \end{aligned} \quad (3.20)$$

By (2.13), from (3.20) we conclude that relation (3.19) follows if

$$\max_{1 \leq k \leq KNn} \left| \sum_{i=2}^{k-1} f'\left(\frac{1}{\sqrt{n}} \sum_{j=1}^{i-1} u_j\right) \left(\frac{T_i}{n} - \frac{T_{i-1}}{n} - \frac{\sigma_\epsilon^2}{n}\right) \right| \rightarrow_P 0 \quad (3.21)$$

and

$$\max_{1 \leq k \leq KNn} \left| f'\left(\frac{1}{\sqrt{n}} \sum_{j=1}^{k-1} u_j\right) \right| \left| \frac{T_{k-1}}{n} - \frac{k-1}{n} \sigma_\epsilon^2 \right| \rightarrow_P 0. \quad (3.22)$$

By Lemma 11.1 and estimate (12.12), under the assumptions of the theorem,

$$\eta_{tn} = f'\left(\frac{1}{\sqrt{n}} \sum_{j=1}^{t-1} u_j\right) (T_t - T_{t-1} - \sigma_\epsilon^2),$$

$t \geq 2$ , is a martingale-difference sequence with

$$\max_{1 \leq t \leq n} E\eta_{tn}^2 \leq L_1 E\epsilon_0^4 \max_{1 \leq t \leq n} E\left(f'\left(\frac{1}{\sqrt{n}} \sum_{j=1}^{t-1} u_j\right)\right)^2 \leq L_2$$

for some constants  $L_1 > 0$  and  $L_2 > 0$ . Therefore, from Lemma 12.5 we conclude that (3.21) holds. In addition, from Theorem 2.2 it follows that

$$\max_{1 \leq k \leq KNn} \left| f'\left(\frac{1}{\sqrt{n}} \sum_{j=1}^{k-1} u_j\right) \right| = O_P(1). \quad (3.23)$$

This, together with (11.3), implies (3.22). Consequently, (3.19) indeed holds.

By definition of  $B(s, \alpha)$  and  $C(s, \alpha)$  in (10.7) with  $g_1(x) = f(C(1)x)$  and  $g_2(x) = h_0(1)f'(C(1)x)$ , we have that

$$B(s, X_n) = \left( \int_0^s h_0(1)f'(C(1)W(v))dv, 0 \right) = (\tilde{B}^1(s), \tilde{B}^2(s)), \quad (3.24)$$

where  $\tilde{B}^1(s) = \int_0^s h_0(1)f'(C(1)W(v))dv$  and  $\tilde{B}^2(s) = 0$ , and

$$C(s, X_n) = \begin{pmatrix} \int_0^s f^2(C(1)W(v))dv & \int_0^s f(C(1)W(v))dv \\ \int_0^s f(C(1)W(v))dv & s \end{pmatrix} = \begin{pmatrix} \tilde{C}^{11}(s) & \tilde{C}^{12}(s) \\ \tilde{C}^{21}(s) & \tilde{C}^{22}(s) \end{pmatrix}. \quad (3.25)$$

By (3.18) and (3.24), for  $\frac{T_{k-1}}{n} < s \leq \frac{T_k}{n}$ ,  $k = 1, 2, \dots$ ,

$$\begin{aligned} |\tilde{B}_n^1(s) - \tilde{B}^1(s)| &= |h_0(1)| \left| \sum_{i=1}^{k-1} \int_{\frac{T_{i-1}}{n}}^{\frac{T_i}{n}} \left[ f' \left( \frac{1}{\sqrt{n}} \sum_{j=1}^{i-1} u_j \right) - f'(C(1)W(v)) \right] dv \right. \\ &\quad \left. + \int_{\frac{T_{k-1}}{n}}^s \left[ f' \left( \frac{1}{\sqrt{n}} \sum_{j=1}^{k-1} u_j \right) - f'(C(1)W(v)) \right] dv \right| \\ &\leq s|h_0(1)| \max_{1 \leq i \leq k} \sup_{v \in [\frac{T_{i-1}}{n}, \frac{T_i}{n}]} \left| f' \left( \frac{1}{\sqrt{n}} \sum_{j=1}^{i-1} u_j \right) - f'(C(1)W(v)) \right|. \end{aligned} \quad (3.26)$$

Thus, for  $\frac{T_{k-1}}{n} < N \leq \frac{T_k}{n}$ ,  $k = 1, 2, \dots$ ,

$$\begin{aligned} \sup_{0 \leq s \leq N} |\tilde{B}_n^1(s) - \tilde{B}^1(s)| &\leq \\ N|h_0(1)| \max_{1 \leq i \leq k} \sup_{v \in [\frac{T_{i-1}}{n}, \frac{T_i}{n}]} \left| f' \left( \frac{1}{\sqrt{n}} \sum_{j=1}^{i-1} u_j \right) - f'(C(1)W(v)) \right|. \end{aligned} \quad (3.27)$$

By (11.1) we have

$$\begin{aligned} \max_{1 \leq i \leq k} \sup_{v \in [\frac{T_{i-1}}{n}, \frac{T_i}{n}]} \left| f' \left( \frac{1}{\sqrt{n}} \sum_{j=1}^{i-1} u_j \right) - f'(C(1)W(v)) \right| &\leq \\ \max_{1 \leq i \leq k} \left| f' \left( \frac{1}{\sqrt{n}} \sum_{j=1}^{i-1} u_j \right) - f' \left( C(1)W \left( \frac{T_{i-1}}{n} \right) \right) \right| &+ \end{aligned}$$

$$\begin{aligned}
& \max_{1 \leq i \leq k} \sup_{v \in [\frac{T_{i-1}}{n}, \frac{T_i}{n}]} \left| f' \left( C(1)W \left( \frac{T_{i-1}}{n} \right) \right) - f'(C(1)W(v)) \right| \leq \\
& \max_{1 \leq i \leq k} \left| f' \left( \frac{1}{\sqrt{n}} \sum_{j=1}^{i-1} u_j \right) - f' \left( \frac{C(1)}{\sqrt{n}} \sum_{j=1}^{i-1} \epsilon_j \right) \right| + \\
& \max_{1 \leq i \leq k} \sup_{v \in [\frac{T_{i-1}}{n}, \frac{T_i}{n}]} \left| f' \left( C(1)W \left( \frac{T_{i-1}}{n} \right) \right) - f'(C(1)W(v)) \right| \leq \\
& \max_{1 \leq i \leq k} \left| f' \left( \frac{1}{\sqrt{n}} \sum_{j=1}^{i-1} u_j \right) - f' \left( \frac{C(1)}{\sqrt{n}} \sum_{j=1}^{i-1} \epsilon_j \right) \right| + \\
& \max_{1 \leq i \leq k} \sup_{v_1, v_2 \in [\frac{T_{i-1}}{n}, \frac{T_i}{n}]} \left| f'(C(1)W(v_1)) - f'(C(1)W(v_2)) \right|. \tag{3.28}
\end{aligned}$$

Using (2.6) we get

$$\begin{aligned}
& \max_{1 \leq i \leq KNn} \left| f' \left( \frac{1}{\sqrt{n}} \sum_{j=1}^{i-1} u_j \right) - f' \left( \frac{C(1)}{\sqrt{n}} \sum_{j=1}^{i-1} \epsilon_j \right) \right| \\
& = \max_{1 \leq i \leq KNn} \left| f' \left( \frac{C(1)}{\sqrt{n}} \sum_{j=1}^{i-1} \epsilon_j + \frac{\tilde{\epsilon}_0}{\sqrt{n}} - \frac{\tilde{\epsilon}_{i-1}}{\sqrt{n}} \right) - f' \left( \frac{C(1)}{\sqrt{n}} \sum_{j=1}^{i-1} \epsilon_j \right) \right|. \tag{3.29}
\end{aligned}$$

By (2.8), from (3.29) and uniform continuity of  $f'$  on compacts we obtain that

$$\max_{1 \leq i \leq KNn} \left| f' \left( \frac{1}{\sqrt{n}} \sum_{j=1}^{i-1} u_j \right) - f' \left( \frac{C(1)}{\sqrt{n}} \sum_{j=1}^{i-1} \epsilon_j \right) \right| \rightarrow_P 0. \tag{3.30}$$

In addition, relation (2.16), together with uniform continuity of  $f'$  on compacts and that of the Brownian sample paths, implies

$$\max_{1 \leq i \leq KNn} \sup_{v_1, v_2 \in [\frac{T_{i-1}}{n}, \frac{T_i}{n}]} \left| f'(C(1)W(v_1)) - f'(C(1)W(v_2)) \right| \rightarrow_P 0, \tag{3.31}$$

By (2.13), from (3.27), (3.28), (3.30) and (3.31) we get

$$\sup_{0 \leq s \leq N} |\tilde{B}_n^1(s) - \tilde{B}^1(s)| \rightarrow_P 0 \tag{3.32}$$

for all  $N \in \mathbf{N}$ . From (3.19) and (3.32) we conclude that

$$\sup_{0 \leq s \leq N} |B_n^1(s) - \tilde{B}^1(s)| \rightarrow_P 0. \tag{3.33}$$

Consequently, condition  $[sup - \beta]$  (and thus  $[sup - \beta_{loc}]$ ) of Theorem 9.1 is satisfied.

By (3.12), (3.13) and (3.25), for  $\frac{T_{k-1}}{n} < s \leq \frac{T_k}{n}$ ,  $k = 1, 2, \dots$ ,

$$\begin{aligned}
|C_n^{11}(s) - \tilde{C}^{11}(s)| &= \left| \sum_{i=1}^{k-1} \int_{\frac{T_{i-1}}{n}}^{\frac{T_i}{n}} \left[ f^2\left(\frac{1}{\sqrt{n}} \sum_{j=1}^{i-1} u_j\right) - f^2(C(1)W(v)) \right] dv \right. \\
&\quad \left. + \int_{\frac{T_{k-1}}{n}}^s \left[ f^2\left(\frac{1}{\sqrt{n}} \sum_{j=1}^{k-1} u_j\right) - f^2(C(1)W(v)) \right] dv \right| \\
&\leq s \max_{1 \leq i \leq k} \sup_{v \in [\frac{T_{i-1}}{n}, \frac{T_i}{n}]} \left| f^2\left(\frac{1}{\sqrt{n}} \sum_{j=1}^{i-1} u_j\right) - f^2(C(1)W(v)) \right|, \tag{3.34}
\end{aligned}$$

$$\begin{aligned}
|C_n^{12}(s) - \tilde{C}^{12}(s)| &= |C_n^{21}(s) - \tilde{C}^{21}(s)| = \\
&= \left| \sum_{i=1}^{k-1} \int_{\frac{T_{i-1}}{n}}^{\frac{T_i}{n}} \left[ f\left(\frac{1}{\sqrt{n}} \sum_{j=1}^{i-1} u_j\right) - f(C(1)W(v)) \right] dv \right. \\
&\quad \left. + \int_{\frac{T_{k-1}}{n}}^s \left[ f\left(\frac{1}{\sqrt{n}} \sum_{j=1}^{k-1} u_j\right) - f(C(1)W(v)) \right] dv \right| \\
&\leq s \max_{1 \leq i \leq k} \sup_{v \in [\frac{T_{i-1}}{n}, \frac{T_i}{n}]} \left| f\left(\frac{1}{\sqrt{n}} \sum_{j=1}^{i-1} u_j\right) - f(C(1)W(v)) \right|. \tag{3.35}
\end{aligned}$$

Thus, for  $\frac{T_{k-1}}{n} < N \leq \frac{T_k}{n}$ ,  $k = 1, 2, \dots$ ,

$$\begin{aligned}
&\sup_{0 \leq s \leq N} |C_n^{11}(s) - \tilde{C}^{11}(s)| \leq \\
&N \max_{1 \leq i \leq k} \sup_{v \in [\frac{T_{i-1}}{n}, \frac{T_i}{n}]} \left| f^2\left(\frac{1}{\sqrt{n}} \sum_{j=1}^{i-1} u_j\right) - f^2(C(1)W(v)) \right|, \tag{3.36}
\end{aligned}$$

$$\begin{aligned}
&\sup_{0 \leq s \leq N} |C_n^{12}(s) - \tilde{C}^{12}(s)| \leq \\
&N \max_{1 \leq i \leq k} \sup_{v \in [\frac{T_{i-1}}{n}, \frac{T_i}{n}]} \left| f\left(\frac{1}{\sqrt{n}} \sum_{j=1}^{i-1} u_j\right) - f(C(1)W(v)) \right|. \tag{3.37}
\end{aligned}$$

By (11.1) and similar to (3.28), we have

$$\max_{1 \leq i \leq k} \sup_{v \in [\frac{T_{i-1}}{n}, \frac{T_i}{n}]} \left| f^2\left(\frac{1}{\sqrt{n}} \sum_{j=1}^{i-1} u_j\right) - f^2(C(1)W(v)) \right| \leq$$

$$\begin{aligned}
& \max_{1 \leq i \leq k} \left| f^2 \left( \frac{1}{\sqrt{n}} \sum_{j=1}^{i-1} u_j \right) - f^2 \left( \frac{C(1)}{\sqrt{n}} \sum_{j=1}^{i-1} \epsilon_j \right) \right| + \\
& \max_{1 \leq i \leq k} \sup_{v_1, v_2 \in [\frac{T_{i-1}}{n}, \frac{T_i}{n}]} \left| f^2(C(1)W(v_1)) - f^2(C(1)W(v_2)) \right|
\end{aligned} \tag{3.38}$$

and

$$\begin{aligned}
& \max_{1 \leq i \leq k} \sup_{v \in [\frac{T_{i-1}}{n}, \frac{T_i}{n}]} \left| f \left( \frac{1}{\sqrt{n}} \sum_{j=1}^{i-1} u_j \right) - f(C(1)W(v)) \right| \leq \\
& \max_{1 \leq i \leq k} \left| f \left( \frac{1}{\sqrt{n}} \sum_{j=1}^{i-1} u_j \right) - f \left( \frac{C(1)}{\sqrt{n}} \sum_{j=1}^{i-1} \epsilon_j \right) \right| + \\
& \max_{1 \leq i \leq k} \sup_{v_1, v_2 \in [\frac{T_{i-1}}{n}, \frac{T_i}{n}]} \left| f(C(1)W(v_1)) - f(C(1)W(v_2)) \right|.
\end{aligned} \tag{3.39}$$

By (2.6) we have

$$\begin{aligned}
& \max_{1 \leq i \leq KNn} \left| f^2 \left( \frac{1}{\sqrt{n}} \sum_{j=1}^{i-1} u_j \right) - f^2 \left( \frac{C(1)}{\sqrt{n}} \sum_{j=1}^{i-1} \epsilon_j \right) \right| \\
& = \max_{1 \leq i \leq KNn} \left| f^2 \left( \frac{C(1)}{\sqrt{n}} \sum_{j=1}^{i-1} \epsilon_j + \tilde{\epsilon}_0 - \tilde{\epsilon}_{i-1} \right) - f^2 \left( \frac{C(1)}{\sqrt{n}} \sum_{j=1}^{i-1} \epsilon_j \right) \right|,
\end{aligned} \tag{3.40}$$

$$\begin{aligned}
& \max_{1 \leq i \leq KNn} \left| f \left( \frac{1}{\sqrt{n}} \sum_{j=1}^{i-1} u_j \right) - f \left( \frac{C(1)}{\sqrt{n}} \sum_{j=1}^{i-1} \epsilon_j \right) \right| \\
& = \max_{1 \leq i \leq KNn} \left| f \left( \frac{C(1)}{\sqrt{n}} \sum_{j=1}^{i-1} \epsilon_j + \tilde{\epsilon}_0 - \tilde{\epsilon}_{i-1} \right) - f \left( \frac{C(1)}{\sqrt{n}} \sum_{j=1}^{i-1} \epsilon_j \right) \right|.
\end{aligned} \tag{3.41}$$

By (2.8), from (3.40) and (3.41) and uniform continuity of  $f$  and  $f^2$  on compacts we obtain

$$\max_{1 \leq i \leq KNn} \left| f^2 \left( \frac{1}{\sqrt{n}} \sum_{j=1}^{i-1} u_j \right) - f^2 \left( \frac{C(1)}{\sqrt{n}} \sum_{j=1}^{i-1} \epsilon_j \right) \right| \rightarrow_P 0, \tag{3.42}$$

$$\max_{1 \leq i \leq KNn} \left| f \left( \frac{1}{\sqrt{n}} \sum_{j=1}^{i-1} u_j \right) - f \left( \frac{C(1)}{\sqrt{n}} \sum_{j=1}^{i-1} \epsilon_j \right) \right| \rightarrow_P 0. \tag{3.43}$$



In addition, relation (2.16), together with uniform continuity of  $f$  and  $f^2$  on compacts and of the Brownian sample paths implies that

$$\max_{1 \leq i \leq KNn} \sup_{v_1, v_2 \in [\frac{T_{i-1}}{n}, \frac{T_i}{n}]} \left| f^2(C(1)W(v_1)) - f^2(C(1)W(v_2)) \right| \rightarrow_P 0, \quad (3.44)$$

$$\max_{1 \leq i \leq KNn} \sup_{v_1, v_2 \in [\frac{T_{i-1}}{n}, \frac{T_i}{n}]} \left| f(C(1)W(v_1)) - f(C(1)W(v_2)) \right| \rightarrow_P 0, \quad (3.45)$$

By (2.13), from (3.36)-(3.39) and (3.42)-(3.45) we get

$$\sup_{0 \leq s \leq N} |C_n^{11}(s) - \tilde{C}^{11}(s)| \rightarrow_P 0, \quad (3.46)$$

$$\sup_{0 \leq s \leq N} |C_n^{12}(s) - \tilde{C}^{12}(s)| = \sup_{0 \leq s \leq N} |C_n^{21}(s) - \tilde{C}^{21}(s)| \rightarrow_P 0, \quad (3.47)$$

for all  $N \in \mathbf{N}$ . Relations (3.46) and (3.47), together with  $C_n^{22}(s) = \tilde{C}^{22}(s) = s$  evidently imply that

$$\sup_{0 \leq s \leq N} |C_n(s) - C(s, X_n)| \rightarrow_P 0,$$

for all  $N \in \mathbf{N}$ . Consequently, condition  $[sup - \gamma]$  (and thus  $[\gamma_{loc} - \mathbf{R}_+^2]$ ) of Theorem 9.1 is satisfied. We therefore have  $X_n \rightarrow_d X$ . This, together with (2.3) and (3.8) implies, by Lemma 12.2, relation (3.6).

For  $k \geq 2$ , denote

$$\begin{aligned} \mathcal{I}_k &= \left| \frac{1}{\sqrt{n}} \sum_{t=2}^k f\left(\frac{1}{\sqrt{n}} \sum_{i=1}^{t-1} u_i\right) u_t - \right. \\ &\quad \left. \frac{\lambda}{n} \sum_{t=2}^k f'\left(\frac{1}{\sqrt{n}} \sum_{i=1}^{t-1} u_i\right) - \frac{C(1)}{\sqrt{n}} \sum_{t=2}^k f\left(\frac{1}{\sqrt{n}} \sum_{i=1}^{t-1} u_i\right) \epsilon_t \right|. \end{aligned}$$

To complete the proof, we show that, for all  $N \in \mathbf{N}$ ,

$$\sup_{0 \leq r \leq N} \mathcal{I}_{[nr]} \rightarrow_P 0. \quad (3.48)$$

Using (2.5) and summation by parts gives

$$\begin{aligned} \mathcal{I}_k &= \left| \frac{1}{\sqrt{n}} \sum_{t=2}^k f\left(\frac{1}{\sqrt{n}} \sum_{i=1}^{t-1} u_i\right) (\tilde{\epsilon}_{t-1} - \tilde{\epsilon}_t) - \frac{\lambda}{n} \sum_{t=2}^k f'\left(\frac{1}{\sqrt{n}} \sum_{i=1}^{t-1} u_i\right) \right| = \\ &= \left| -\frac{1}{\sqrt{n}} f\left(\frac{1}{\sqrt{n}} \sum_{i=1}^k u_i\right) \tilde{\epsilon}_k + \frac{1}{\sqrt{n}} \sum_{t=2}^k \left( f\left(\frac{1}{\sqrt{n}} \sum_{i=1}^t u_i\right) - f\left(\frac{1}{\sqrt{n}} \sum_{i=1}^{t-1} u_i\right) \right) \tilde{\epsilon}_t - \right. \\ &\quad \left. \frac{\lambda}{n} \sum_{t=2}^k f'\left(\frac{1}{\sqrt{n}} \sum_{i=1}^{t-1} u_i\right) \right|. \end{aligned}$$

Consequently, for all  $N \in \mathbf{N}$ ,

$$\begin{aligned}
\max_{1 \leq k \leq nN} \mathcal{I}_k &\leq \max_{1 \leq k \leq nN} \left| \frac{1}{\sqrt{n}} f \left( \frac{1}{\sqrt{n}} \sum_{i=1}^k u_i \right) \tilde{\epsilon}_k \right| + \\
&\max_{1 \leq k \leq nN} \left| \frac{1}{n} \sum_{t=2}^k f' \left( \frac{1}{\sqrt{n}} \sum_{i=1}^{t-1} u_i \right) (u_t \tilde{\epsilon}_t - \lambda) \right| + \\
&\max_{1 \leq k \leq nN} \left| \frac{1}{\sqrt{n}} \sum_{t=2}^k \left( f \left( \frac{1}{\sqrt{n}} \sum_{i=1}^t u_i \right) - f \left( \frac{1}{\sqrt{n}} \sum_{i=1}^{t-1} u_i \right) - \right. \right. \\
&\left. \left. f' \left( \frac{1}{\sqrt{n}} \sum_{i=1}^{t-1} u_i \right) \frac{u_t}{\sqrt{n}} \right) \tilde{\epsilon}_t \right| = \mathcal{I}_{1n} + \mathcal{I}_{2n} + \mathcal{I}_{3n}.
\end{aligned} \tag{3.49}$$

From (2.8) and property (3.23) it follows that  $\mathcal{I}_{1n} \rightarrow_P 0$ .

Similar to the derivations of second order BN decompositions in Phillips and Solo (1992) and the proof of Theorem 2.4, it is not difficult to see that

$$\begin{aligned}
u_t \tilde{\epsilon}_t &= h_0(L) \epsilon_t^2 + \sum_{r=1}^{\infty} h_r(L) \epsilon_t \epsilon_{t-r} = \\
&h_0(1) \epsilon_t^2 - (1-L) \tilde{w}_{at} + \epsilon_t \epsilon_{t-1}^h - (1-L) \tilde{w}_{bt},
\end{aligned} \tag{3.50}$$

where  $\tilde{w}_{at} = \tilde{h}_0(L) \epsilon_t^2$ ,  $\epsilon_{t-1}^h = \sum_{r=1}^{\infty} h_r(1) \epsilon_{t-r}$  and  $\tilde{w}_{bt} = \sum_{r=1}^{\infty} \tilde{h}_r(L) \epsilon_t \epsilon_{t-r}$  (the validity of decomposition (3.50) is justified by Lemma 12.9).

Using (3.9) and (3.50), we get that

$$\begin{aligned}
\mathcal{I}_{2n} &\leq \max_{1 \leq k \leq nN} \left| \frac{1}{n} \sum_{t=2}^k f' \left( \frac{1}{\sqrt{n}} \sum_{i=1}^{t-1} u_i \right) (h_0(1) \epsilon_t^2 - h_0(1) \sigma_\epsilon^2) \right| + \\
&\max_{1 \leq k \leq nN} \left| \frac{1}{n} \sum_{t=2}^k f' \left( \frac{1}{\sqrt{n}} \sum_{i=1}^{t-1} u_i \right) \epsilon_t \epsilon_{t-1}^h \right| + \\
&\max_{1 \leq k \leq nN} \left| \frac{1}{n} \sum_{t=2}^k f' \left( \frac{1}{\sqrt{n}} \sum_{i=1}^{t-1} u_i \right) (\tilde{w}_{at} - \tilde{w}_{a,t-1}) \right| + \\
&\max_{1 \leq k \leq nN} \left| \frac{1}{n} \sum_{t=2}^k f' \left( \frac{1}{\sqrt{n}} \sum_{i=1}^{t-1} u_i \right) (\tilde{w}_{bt} - \tilde{w}_{b,t-1}) \right| =
\end{aligned}$$

$$= \mathcal{I}_{2n}^{(1)} + \mathcal{I}_{2n}^{(2)} + \mathcal{I}_{2n}^{(3)} + \mathcal{I}_{2n}^{(4)}.$$

As in the proof of Theorem 3.1 and relation (3.19) above, we conclude, by Lemma 12.12, that  $\eta_{tn}^{(1)} = f'\left(\frac{1}{\sqrt{n}} \sum_{i=1}^{t-1} u_i\right)(\epsilon_t^2 - \sigma_\epsilon^2)$ ,  $t \geq 2$ , is a martingale-difference with

$$\max_{1 \leq t \leq n} E\left(\eta_{tn}^{(1)}\right) \leq L_1 E\epsilon_0^4 \max_{1 \leq t \leq n} E\left(f'\left(\frac{1}{\sqrt{n}} \sum_{i=1}^{t-1} u_i\right)\right)^2 \leq L_2$$

for some constants  $L_1 > 0$  and  $L_2 > 0$ .

Similarly, from Lemmas 12.12 and 12.11 it follows, by Hölder's inequality, that the martingale-difference sequence  $\eta_{tn}^{(2)} = f'\left(\frac{1}{\sqrt{n}} \sum_{i=1}^{t-1} u_i\right)\eta_t \eta_{t-1}^h$ ,  $t \geq 2$ , satisfies

$$\max_{1 \leq t \leq nN} E\left(\eta_{tn}^{(2)}\right) = E\epsilon_0^2 \max_{1 \leq t \leq nN} E\left(f'\left(\frac{1}{\sqrt{n}} \sum_{i=1}^{t-1} u_i\right)\right)^2 (\eta_{t-1}^h)^2 \leq$$

$$E\epsilon_0^2 \left[E(\eta_{t-1}^h)^4\right]^{1/2} \max_{1 \leq t \leq nN} \left[E\left(f'\left(\frac{1}{\sqrt{n}} \sum_{i=1}^{t-1} u_i\right)\right)^4\right]^{1/2} \leq L$$

for some constant  $L > 0$ . Using Theorem 12.5, we, therefore, have

$$\mathcal{I}_{2n}^{(1)} = \max_{1 \leq k \leq nN} \left| \frac{1}{n} \sum_{t=2}^k \eta_{tn}^{(1)} \right| \rightarrow_P 0$$

and

$$\mathcal{I}_{2n}^{(2)} \max_{1 \leq k \leq nN} \left| \frac{1}{n} \sum_{t=2}^k \eta_{tn}^{(2)} \right| \rightarrow_P 0.$$

In addition, using summation by parts and the smoothness assumptions on  $f$ , we find that (below,  $S_k = \sum_{i=1}^k u_i$ )

$$\begin{aligned} \mathcal{I}_{2n}^{(3)} &\leq \max_{1 \leq k \leq nN} \left| \frac{1}{n} f'\left(\frac{1}{\sqrt{n}} \sum_{i=1}^k u_i\right) \tilde{w}_{ak} \right| + \\ &\max_{1 \leq k \leq nN} \left| \frac{1}{n} \sum_{t=2}^k \left( f'\left(\frac{1}{\sqrt{n}} \sum_{i=1}^t u_i\right) - f'\left(\frac{1}{\sqrt{n}} \sum_{i=1}^{t-1} u_i\right) \right) \tilde{w}_{at} \right| \leq \\ &\max_{1 \leq k \leq nN} \left| \frac{1}{n} f'\left(\frac{1}{\sqrt{n}} \sum_{i=1}^k u_i\right) \right| \max_{1 \leq k \leq nN} \frac{1}{n} |\tilde{w}_{ak}| + \\ &N \max_{1 \leq k \leq nN} \frac{1}{\sqrt{n}} |u_k \tilde{w}_{ak}| \sup_{|t| \leq \max_{0 \leq k \leq nN} |S_k|/\sqrt{n}} |f''(t)|, \end{aligned} \tag{3.51}$$

$$\begin{aligned}
\mathcal{I}_{2n}^{(4)} &\leq \max_{1 \leq k \leq nN} \left| \frac{1}{n} f' \left( \frac{1}{\sqrt{n}} \sum_{i=1}^k u_i \right) \tilde{w}_{bk} \right| + \\
&\max_{1 \leq k \leq nN} \left| \frac{1}{n} \sum_{t=2}^k \left( f' \left( \frac{1}{\sqrt{n}} \sum_{i=1}^t u_i \right) - f' \left( \frac{1}{\sqrt{n}} \sum_{i=1}^{t-1} u_i \right) \right) \tilde{w}_{bt} \right| \leq \\
&\max_{1 \leq k \leq nN} \left| \frac{1}{n} f' \left( \frac{1}{\sqrt{n}} \sum_{i=1}^k u_i \right) \right| \max_{1 \leq k \leq nN} \frac{1}{n} |\tilde{w}_{bk}| + \\
&N \max_{1 \leq k \leq nN} \frac{1}{\sqrt{n}} |u_k \tilde{w}_{bk}| \sup_{|t| \leq \max_{0 \leq k \leq nN} |S_k|/\sqrt{n}} |f''(t)|.
\end{aligned} \tag{3.52}$$

By Lemma 12.10,  $\sup_t |\tilde{w}_{at}|^3 \rightarrow_P 0$  and  $\sup_t |\tilde{w}_{bt}|^3 \rightarrow_P 0$  under the assumptions of the theorem. Therefore, using Lemma 12.4 with  $p = 6$  we have

$$\max_{1 \leq k \leq nN} n^{-1/6} |u_k| \rightarrow_P 0, \quad \max_{1 \leq k \leq nN} n^{-1/3} |\tilde{w}_{ak}| \rightarrow_P 0, \tag{3.53}$$

$$\max_{1 \leq k \leq nN} n^{-1/3} |\tilde{w}_{bk}| \rightarrow_P 0.$$

These relations also imply that  $\max_{1 \leq k \leq nN} n^{-1/2} |u_k \tilde{w}_{ak}| \rightarrow_P 0$  and  $\max_{1 \leq k \leq nN} n^{-1/2} |u_k \tilde{w}_{bk}| \rightarrow_P 0$ . By Theorem 2.2,

$$\max_{1 \leq k \leq nN} n^{-1/2} \left| \sum_{t=1}^k u_t \right| = O_P(1). \tag{3.54}$$

The above, together with (3.23), (3.51) and (3.52), we conclude that  $\mathcal{I}_{2n}^{(3)} \rightarrow_P 0$  and  $\mathcal{I}_{2n}^{(4)} \rightarrow_P 0$ .

We have, by Taylor expansion, that

$$\begin{aligned}
&\max_{0 \leq k \leq nN} \left| \frac{1}{\sqrt{n}} \sum_{t=2}^k \left( f \left( \frac{1}{\sqrt{n}} \sum_{i=1}^t u_i \right) - f \left( \frac{1}{\sqrt{n}} \sum_{i=1}^{t-1} u_i \right) - f' \left( \frac{1}{\sqrt{n}} \sum_{i=1}^{t-1} u_i \right) \frac{u_t}{\sqrt{n}} \right) \tilde{\epsilon}_t \right| \\
&\leq (N/2) \max_{1 \leq k \leq nN} \frac{1}{\sqrt{n}} u_k^2 |\tilde{\epsilon}_k| \sup_{|t| \leq \max_{0 \leq k \leq nN} |S_k|/\sqrt{n}} |f''(t)|.
\end{aligned} \tag{3.55}$$

By Lemmas 12.4 and 12.10,  $\max_{1 \leq k \leq nN} n^{-1/6} |\tilde{\epsilon}_k| \rightarrow_P 0$ . This, together with (3.54) and the first relation in (3.53) leads to  $\max_{0 \leq k \leq nN} n^{-1/2} u_k^2 |\tilde{\epsilon}_k| \rightarrow_P 0$ . Consequently, by (3.55) we have  $\mathcal{I}_{3n} \rightarrow_P 0$ .

From (3.49) we deduce that (3.48) indeed holds. By Lemmas 12.1 and 12.3, relations (3.6) and (3.48) imply (3.1). ■

## 4 Useful multivariate extensions

The present section shows how to skip the Skorohod embedding argument at the beginning of the proofs, which is used above to convert discrete time martingales and semimartingales to continuous versions (e.g. in (2.2), (2.11) and (3.8)) and thereby simplify some of the arguments. The approach is to work directly with the discrete time processes as discontinuous processes and seek to verify conditions for martingale and semimartingale convergence that involve the predictable measures of jumps for the discontinuous processes. This may be accomplished by using suitable additional conditions beyond those we have already employed in Theorems 9.1 and 9.2. Dealing with these additional conditions is not problematic, and the increase in the technical difficulty is justified in view of the wide range of applications covered by these more general results. The extensions include results on convergence to multivariate stochastic integrals and a precise formulation of the unification theorem for stationary and nonstationary autoregression. To simplify presentation of the results, we treat the bivariate case here and extensions to general multivariate cases follow in the same fashion.

We start with the following martingale convergence result, which provides a limit theory for multivariate stochastic integrals and enables later extension to the case of general linear processes.

The argument for the results in this section relies on application of Theorem IX.3.25 in JS that gives conditions for convergence of general (not necessarily continuous) square integrable semimartingales  $X_n$  in terms of their first characteristics without truncation,  $B'_n$ , second modified characteristics without truncation,  $\tilde{C}'_n$  and the predictable measures of jumps,  $\nu_n$ , defined in JS, Ch. II, §2 and IX.3.25. While the formulations of definitions of these concepts in the general case are quite cumbersome, they simplify when the semimartingales of interest are continuous-time analogues of respective discrete-time processes, as in most of the econometric models encountered in practice.

Let  $(Y_n(k))_{k=0}^\infty$ ,  $Y_n(k) = (Y_n^1(k), \dots, Y_n^d(k))$ ,  $k = 0, 1, 2, \dots$ , be a sequence of discrete-time semimartingales on the space  $(\Omega, \mathfrak{F}, P)$  with the filtration  $\mathfrak{F}_0 = (\Omega, \emptyset) \subseteq \mathfrak{F}_1 \subseteq \dots \subseteq \mathfrak{F}$ :

$$Y_n^j(k) = \sum_{t=0}^k \eta_n^j(t) = \eta_n^j(0) + \sum_{t=1}^k m_n^j(t) + \sum_{t=1}^k b_n^j(t),$$

$j = 1, 2, \dots, d$ , where  $\eta_n^j(t) = Y_n^j(t) - Y_n^j(t-1)$ ,  $t \geq 1$ , and  $m_n^j(t) = \eta_n^j(t) - E(\eta_n^j(t)|\mathfrak{F}_{t-1})$  and  $b_n^j(t) = E(\eta_n^j(t)|\mathfrak{F}_{t-1})$ ,  $t \geq 1$ , are, respectively, the components of the martingale and predictable part in the discrete-time analogue of representation (8.1).

In the case where the sequence  $(X_n(s), s \geq 0)$ ,  $n \geq 1$ , of semimartingales whose convergence is studied is given by continuous-time analogues of discrete-time processes  $Y_n$  defined by  $X_n(s) = Y_n([ns])$ ,  $s \geq 0$ , the modified characteristics of  $X_n$  are given by similar continuous-time analogues of predictable characteristics of  $Y_n$ .

Namely, the first modified characteristic of  $X_n$  is the  $\mathbf{R}^d$ -valued process  $(B'_n(s), s \geq 0)$ ,  $B'_n(s) = (\tilde{B}_n^1(s), \dots, \tilde{B}_n^d(s))$ , where  $\tilde{B}_n^j(s) = \sum_{t=1}^{[ns]} b_n^j(t)$ , and the second modified characteristic of  $X_n$  is the process  $(C'_n(s), s \geq 0)$ ,  $\tilde{C}'_n(s) = (\tilde{C}_n^{ij}(s))_{1 \leq i, j \leq d}$ , where  $\tilde{C}_n^{ij}(s) = \sum_{t=1}^{[ns]} E[m_n^i(t)m_n^j(t)|\mathfrak{F}_{t-1}]$ . In addition to that, one has the following representation for the integral of a continuous function  $g$  on  $\mathbf{R}^d$  with respect to the measure  $\nu_n$  that appears in Theorem IX.3.25 in JS employed in the argument for the results in this section of the paper (provided that the integral and the expectation exist):  $\int_0^s \int_{\mathbf{R}^d} g(x) \nu_n(dw, dx) = \sum_{t=1}^{[ns]} E[g(\eta_n^1(t), \dots, \eta_n^d(t))|\mathfrak{F}_{t-1}]$ .

Throughout the rest of the paper,  $I(\cdot)$  stands for the indicator function.

**Theorem 4.1** Let  $\{(\epsilon_t, \eta_t)\}_{t=0}^\infty$  be a sequence of i.i.d. mean-zero random vectors such that  $E\epsilon_0^2 = \sigma_\epsilon^2$ ,  $E\eta_0^2 = \sigma_\eta^2$ ,  $E\epsilon_0\eta_0 = \sigma_{\epsilon\eta}$ ,  $E|\epsilon_0|^p < \infty$  and  $E|\eta_0|^p < \infty$  for some  $p > 4$ . Let  $(W, V) = ((W(s), V(s)), s \geq 0)$  be bivariate Brownian motion with covariance matrix

$$\begin{pmatrix} \sigma_\epsilon^2 & \sigma_{\epsilon\eta} \\ \sigma_{\epsilon\eta} & \sigma_\eta^2 \end{pmatrix}.$$

Then

$$\frac{1}{n} \sum_{t=2}^{[nr]} \left( \sum_{i=1}^{t-1} \epsilon_i \right) \eta_t \rightarrow_d \int_0^r W(v) dV(v). \quad (4.1)$$

**Proof.** For  $n \geq 1$ , let  $X_n = (X_n(s), s \geq 0)$  and  $X = (X(s), s \geq 0)$  be the vector martingales

$$X_n(s) = \left( \frac{1}{n} \sum_{t=2}^{[ns]} \left( \sum_{i=1}^{t-1} \epsilon_i \right) \eta_t, \frac{1}{\sqrt{n}} \sum_{t=1}^{[ns]} \epsilon_t, \frac{1}{\sqrt{n}} \sum_{t=1}^{[ns]} \eta_t \right)$$

and

$$X(s) = \left( \int_0^s W(v) dV(v), W(s), V(s) \right) = (X^1(s), X^2(s), X^3(s)).$$

Let  $B'_n = (B'_n(s), s \geq 0)$  denote the first characteristic without truncation of  $X_n$ , let  $\tilde{C}'_n = (\tilde{C}'_n(s), s \geq 0)$  stand for its modified second characteristic without truncation and let  $\nu_n = (\nu_n(ds, dx))$  denote its predictable measure of jumps (see JS, Ch. II, §2 and IX.3.25). The process  $B'_n$  is identically zero so  $B'_n(s) = (0, 0, 0) \in \mathbf{R}^3$ ,  $s \geq 0$ . For the modified second characteristic without truncation of  $X_n$  we have  $\tilde{C}'_n(s) = (\tilde{C}_n^{ij}(s))_{1 \leq i, j \leq 3}$ , where

$$\tilde{C}_n^{11}(s) = \frac{\sigma_\eta^2}{n^2} \sum_{t=2}^{[ns]} \left( \sum_{i=1}^{t-1} \epsilon_i \right)^2,$$

$$\tilde{C}_n^{12}(s) = \tilde{C}_n^{21}(s) = \frac{\sigma_{\epsilon\eta}}{n^{3/2}} \sum_{t=2}^{[ns]} \left( \sum_{i=1}^{t-1} \epsilon_i \right),$$

$$\tilde{C}_n^{13}(s) = \tilde{C}_n^{31}(s) = \frac{\sigma_\eta^2}{n^{3/2}} \sum_{t=2}^{[ns]} \left( \sum_{i=1}^{t-1} \epsilon_i \right),$$

$$\tilde{C}_n^{22}(s) = \frac{\sigma_\epsilon^2 [ns]}{n},$$

$$\tilde{C}_n^{23}(s) = \tilde{C}_n^{32}(s) = \frac{\sigma_{\epsilon\eta} [ns]}{n},$$

$$\tilde{C}_n^{33}(s) = \frac{\sigma_\eta^2 [ns]}{n}.$$

For an element  $\alpha = (\alpha(s), s \geq 0)$ ,  $\alpha(s) = (\alpha_1(s), \alpha_2(s), \alpha_3(s))$  of the Skorohod space  $\mathbb{D}(\mathbf{R}^3)$  and for a Borel subset  $\Gamma$  of  $\mathbf{R}^3$ , let  $B(s, \alpha) = (0, 0, 0)$ ,

$$C(s, \alpha) = \begin{pmatrix} \sigma_\eta^2 \int_0^s \alpha_2^2(v) dv & \sigma_{\epsilon\eta} \int_0^s \alpha_2(v) dv & \sigma_\eta^2 \int_0^s \alpha_2(v) dv \\ \sigma_{\epsilon\eta} \int_0^s \alpha_2(v) dv & \sigma_\epsilon^2 s & \sigma_{\epsilon\eta} s \\ \sigma_\eta^2 \int_0^s \alpha_2(v) dv & \sigma_{\epsilon\eta} s & \sigma_\eta^2 s \end{pmatrix}, \quad (4.2)$$

and  $\nu([0, s], \Gamma)(\alpha) = 0$ . Further, let  $B(\alpha) = (B(s, \alpha), s \geq 0)$ ,  $C(\alpha) = (C(s, \alpha), s \geq 0)$  and  $\nu(\alpha) = (\nu(ds, dx)(\alpha))$ . The process  $X$  is a solution to the stochastic differential equation

$$\begin{aligned} dX^1(s) &= X^2(s) dV(s); \\ dX^2(s) &= dW(s); \\ dX^3(s) &= dV(s), \end{aligned} \quad (4.3)$$

or, equivalently, to stochastic differential equation (10.1) with  $d = 3$  and  $m = 2$  and functions  $b : \mathbf{R}^3 \rightarrow \mathbf{R}^3$  and  $\sigma : \mathbf{R}^3 \rightarrow \mathbf{R}^{3 \times 2}$  given by  $b(x_1, x_2, x_3) = (0, 0, 0)$  and

$$\sigma(x_1, x_2, x_3) = \begin{pmatrix} \sigma_\eta x_2 & 0 \\ \sigma_{\epsilon\eta}/\sigma_\eta & \sqrt{\sigma_\epsilon^2 \sigma_\eta^2 - \sigma_{\epsilon\eta}^2}/\sigma_\eta \\ \sigma_\eta & 0 \end{pmatrix}. \quad (4.4)$$

According to (10.2), the predictable characteristics of  $X$  are  $B(X)$ ,  $C(X)$  and  $\nu(X)$ , with  $B$ ,  $C$  and  $\nu$  defined as above (so that the first and the third predictable characteristics of  $X$  are identically zero, i.e.,  $B = (0, 0, 0) \in \mathbf{R}^3$  and  $\nu = 0$ ). Since  $X$  is continuous, its predictable triplet without truncation is the same.

For  $a \geq 0$  and an element  $\alpha = (\alpha(s), s \geq 0)$  of the Skorohod space  $\mathbb{D}(\mathbf{R}_+^3)$ , define  $S^a(\alpha)$  and  $S_n^a$  as in (9.1). Let  $\mathbb{C}_1(\mathbf{R}^3)$  denote the set of continuous bounded functions  $g : \mathbf{R}^3 \rightarrow \mathbf{R}$  which are equal to zero in a neighborhood of zero. By Theorem IX.3.48 of JS (see also Remark IX.3.40, Theorem III.2.40 and Lemma IX.4.4 in JS and also the proof of Theorem 2.1 in Coffman, Puhalskii and Reiman, 1998), in order to prove that  $X_n \rightarrow_d X$ , it suffices to check that the following conditions hold in addition to conditions (A1)-(A5) of Theorem 9.1 :

$$(A6a) \quad [\delta_{\text{loc}} - \mathbf{R}_+] \quad \int_0^{s \wedge S_n^a} \int_{\mathbf{R}^3} g(x) \nu_n(dw, dx) \rightarrow_P 0 \text{ for all } s > 0, a > 0 \text{ and } g \in \mathbb{C}_1(\mathbf{R}^3).$$

$$[\text{sup} - \beta'_{\text{loc}}] \quad \sup_{0 < s \leq N} |B'_n(s \wedge S_n^a) - B(s \wedge S^a, X_n)| \rightarrow_P 0 \text{ for all } N \in \mathbf{N} \text{ and all } a > 0.$$

$$[\gamma'_{\text{loc}} - \mathbf{R}_+] \quad \tilde{C}'_n(s \wedge S_n^a) - C(s \wedge S^a, X_n) \rightarrow_P 0 \text{ for all } s > 0 \text{ and } a > 0.$$

$$(A7) \quad \lim_{b \rightarrow \infty} \overline{\lim}_{n \rightarrow \infty} P \left( \int_0^{s \wedge S_n^a} \int_{\mathbf{R}^3} |x|^2 I(|x| > b) \nu_n(dw, dx) > \epsilon \right) = 0 \text{ for all } s > 0, a > 0 \text{ and } \epsilon > 0.$$

The following is a sufficient condition for  $[\gamma'_{\text{loc}} - \mathbf{R}_+]$  in (A6a):

$$[\text{sup} - \gamma'] \quad \sup_{0 < s \leq N} |\tilde{C}'_n(s) - C(s, X_n)| \rightarrow_P 0 \text{ for all } N \in \mathbf{N}.$$

In addition, from the definition of the class  $\mathbb{C}_1(\mathbf{R}^3)$  and Lemma 5.5.1 in Liptser and Shiryaev (1989) it follows in a similar way to the proof of Theorem 2.1 in Coffman, Puhalskii and Reiman (1998) that the following is a sufficient condition for  $[\delta_{\text{loc}} - \mathbf{R}_+]$  :

$$[\text{sup} - \Delta] \quad \sup_{0 < s \leq N} |\Delta X_n(s)| \rightarrow_P 0 \text{ for all } N \in \mathbf{N}, \text{ where } \Delta X_n(s) = X_n(s) - X_n(s-).$$

Note that since  $X$  is continuous, in the corresponding results in JS,  $\nu = 0$ ,  $B' = B$  and  $\tilde{C}' = C$ .

Conditions (A1)-(A5) of Theorem 9.1 in the present context can be verified in complete similarity to the proof of Theorem 3.3. In particular, conditions (A2) and (A3) follow from the straightforward extension of Corollary 10.1 to the case of a three-dimensional homogenous diffusion driven by two Brownian motions.

Condition  $[sup - \beta']$  (and thus  $[sup - \beta'_{loc}]$ ) is trivially satisfied since  $B'_n(s) = 0$ ,  $s \geq 0$ , and  $B_n(s, X_n) = 0$ ,  $s \geq 0$ .

From formula (4.2) we have that  $C_n(s, X_n) = (\tilde{C}_n^{ij}(s))_{1 \leq i, j \leq 3}$ , where

$$\tilde{C}_n^{11}(s) = \frac{\sigma_\eta^2}{n^2} \sum_{t=2}^{[ns]} \left( \sum_{i=1}^{t-1} \epsilon_i \right)^2 + \frac{\sigma_\eta^2}{n^2} \left( \sum_{i=1}^{[ns]} \epsilon_i \right)^2 (ns - [ns]) =$$

$$\tilde{C}_n^{11}(s) + \frac{\sigma_\eta^2}{n^2} \left( \sum_{i=1}^{[ns]} \epsilon_i \right)^2 (ns - [ns]),$$

$$\tilde{C}_n^{12}(s) = \tilde{C}_n^{21}(s) = \frac{\sigma_{\epsilon\eta}}{n^{3/2}} \sum_{t=2}^{[ns]} \left( \sum_{i=1}^{t-1} \epsilon_i \right) + \frac{\sigma_{\epsilon\eta}}{n^{3/2}} \left( \sum_{i=1}^{[ns]} \epsilon_i \right) (ns - [ns]) =$$

$$\tilde{C}_n^{12}(s) + \frac{\sigma_{\epsilon\eta}}{n^{3/2}} \left( \sum_{i=1}^{[ns]} \epsilon_i \right) (ns - [ns]),$$

$$\tilde{C}_n^{13}(s) = \tilde{C}_n^{31}(s) = \frac{\sigma_\eta^2}{n^{3/2}} \sum_{t=2}^{[ns]} \left( \sum_{i=1}^{t-1} \epsilon_i \right) + \frac{\sigma_\eta^2}{n^{3/2}} \left( \sum_{i=1}^{[ns]} \epsilon_i \right) (ns - [ns]) =$$

$$\tilde{C}_n^{13}(s) + \frac{\sigma_\eta^2}{n^{3/2}} \left( \sum_{i=1}^{[ns]} \epsilon_i \right) (ns - [ns]),$$

$$\tilde{C}_n^{22}(s) = \sigma_\epsilon^2 s = \tilde{C}_n^{22} + \sigma_\epsilon^2 \frac{ns - [ns]}{n},$$

$$\tilde{C}_n^{23}(s) = \tilde{C}_n^{32}(s) = \sigma_{\epsilon\eta} s = \tilde{C}_n^{23} + \sigma_{\epsilon\eta} \frac{ns - [ns]}{n},$$



$$\tilde{C}_n^{33}(s) = \sigma_\eta^2 s = \tilde{C}_n^{33} + \sigma_\eta^2 \frac{ns - [ns]}{n}.$$

Since, by Lemma 12.5,  $n^{-1} \max_{1 \leq k \leq nN} \left| \sum_{i=1}^k \epsilon_i \right| \rightarrow_P 0$  for all  $N \in \mathbf{N}$ , we thus have

$$\sup_{0 < s \leq N} |\tilde{C}_n^{11}(s) - \tilde{C}_n^{11}(s)| \leq \max_{0 < k \leq nN} \left| \frac{\sigma_\eta^2}{n^2} \left( \sum_{i=1}^k \epsilon_i \right)^2 \right| \rightarrow_P 0,$$

$$\sup_{0 < s \leq N} |\tilde{C}_n^{12}(s) - \tilde{C}_n^{12}(s)| = \sup_{0 < s \leq N} |\tilde{C}_n^{21}(s) - \tilde{C}_n^{21}(s)| \leq$$

$$\max_{0 < k \leq nN} \left| \frac{\sigma_\eta}{n^{3/2}} \left( \sum_{i=1}^k \epsilon_i \right) \right| \rightarrow_P 0,$$

$$\sup_{0 < s \leq N} |\tilde{C}_n^{13}(s) - \tilde{C}_n^{13}(s)| = \sup_{0 < s \leq N} |\tilde{C}_n^{31}(s) - \tilde{C}_n^{31}(s)| \leq$$

$$\max_{0 < k \leq nN} \left| \frac{\sigma_\eta^2}{n^{3/2}} \left( \sum_{i=1}^k \epsilon_i \right) \right| \rightarrow_P 0$$

for all  $N \in \mathbf{N}$ . In addition, evidently,  $\sup_{0 < s \leq N} |\tilde{C}_n^{22}(s) - \tilde{C}_n^{22}(s)| \leq \sigma_\epsilon^2/n \rightarrow_P 0$ ,  $\sup_{0 < s \leq N} |\tilde{C}_n^{23}(s) - \tilde{C}_n^{23}(s)| = \sup_{0 < s \leq N} |\tilde{C}_n^{32}(s) - \tilde{C}_n^{32}(s)| \leq \sigma_{\epsilon\eta}/n \rightarrow_P 0$  and  $\sup_{0 < s \leq N} |\tilde{C}_n^{33}(s) - \tilde{C}_n^{33}(s)| \leq \sigma_\eta^2/n \rightarrow_P 0$  for all  $N \in \mathbf{N}$ . The above obviously implies that  $\sup_{0 < s \leq N} |\tilde{C}'_n(s) - C(s, X_n)| \rightarrow_P 0$  for all  $N \in \mathbf{N}$  and thus condition  $[sup - \gamma']$  (and condition  $[\gamma'_{loc} - \mathbf{R}_+]$ ) is satisfied.

For all  $N \in \mathbf{N}$ , we have

$$\begin{aligned} \sup_{0 \leq s \leq N} |\Delta X_n(s)| &\leq \max_{0 \leq k \leq nN} \frac{1}{\sqrt{n}} \left| \sum_{i=1}^k \epsilon_i \right| \max_{0 \leq k \leq nN} \frac{1}{\sqrt{n}} |\eta_k| + \\ &\max_{0 \leq k \leq nN} \frac{1}{\sqrt{n}} |\epsilon_k| + \max_{0 \leq k \leq nN} \frac{1}{\sqrt{n}} |\eta_k|. \end{aligned}$$

By Theorem 2.1,  $\max_{0 \leq k \leq nN} \frac{1}{\sqrt{n}} \left| \sum_{i=1}^k \epsilon_i \right| = O_P(1)$ . In addition, by Lemma 12.4,  $\max_{0 \leq k \leq nN} \frac{1}{\sqrt{n}} |\epsilon_k| \rightarrow_P 0$  and  $\max_{0 \leq k \leq nN} \frac{1}{\sqrt{n}} |\eta_k| \rightarrow_P 0$ . Using the above, we therefore find that  $\sup_{0 \leq s \leq N} |\Delta X_n(s)| \rightarrow_P 0$  for all  $N \in \mathbf{N}$ . Thus, condition  $[sup - \Delta]$  holds and  $[\delta_{loc} - \mathbf{R}_+]$  holds in consequence.

Finally, we demonstrate that (A7) holds. It is not difficult to see that

$$E \int_0^{s \wedge S_n^a} \int_{\mathbf{R}^3} |x|^2 I(|x| > b) \nu_n(dw, dx) \leq$$

$$\begin{aligned}
& E \int_0^s \int_{\mathbf{R}^3} |x|^2 I(|x| > b) \nu_n(dw, dx) \leq \\
& \frac{1}{b^2} E \int_0^s \int_{\mathbf{R}^3} |x|^4 \nu_n(dw, dx) \leq \\
& \frac{3}{b^2} E \int_0^s \int_{x=(x_1, x_2, x_3) \in \mathbf{R}^3} (x_1^4 + x_2^4 + x_3^4) \nu_n(dw, dx). \tag{4.5}
\end{aligned}$$

Continuing, we have

$$\begin{aligned}
& E \int_0^s \int_{x=(x_1, x_2, x_3) \in \mathbf{R}^3} (x_1^4 + x_2^4 + x_3^4) \nu_n(dw, dx) = \\
& \frac{1}{n^4} \sum_{t=2}^{[ns]} E \left( \sum_{i=1}^{t-1} \epsilon_i \right)^4 E \eta_t^4 + \frac{1}{n^2} \sum_{t=2}^{[ns]} E \epsilon_t^4 + \frac{1}{n^2} \sum_{t=2}^{[ns]} E \eta_t^4 = \\
& \frac{E \eta_0^4}{n^4} \sum_{t=2}^{[ns]} E \left( \sum_{i=1}^{t-1} \epsilon_i \right)^4 + \frac{E \epsilon_0^4 [ns]}{n^2} + \frac{E \eta_0^4 [ns]}{n^2}, \tag{4.6}
\end{aligned}$$

and, using inequality (12.13) in Appendix A5, we find that

$$\frac{E \eta_0^4}{n^4} \sum_{t=2}^{[ns]} E \left( \sum_{i=1}^{t-1} \epsilon_i \right)^4 \leq \frac{K E \epsilon_0^4 E \eta_0^4}{n^2} \sum_{t=2}^{[ns]} t^2 \leq K E \epsilon_0^4 E \eta_0^4 / n \rightarrow 0$$

for all  $s > 0$ . Evidently,  $[ns]/n^2 \rightarrow 0$  for all  $s > 0$ , and from (4.5) and (4.6) we deduce that

$$E \int_0^{s \wedge S_n^a} \int_{\mathbf{R}^3} |x|^2 I(|x| > b) \nu_n(dw, dx) \rightarrow 0$$

for all  $a, b, s > 0$ . By Chebyshev's inequality, this evidently implies that condition (A7) holds.

Consequently, conditions (A1)-(A5) of Theorem 9.1, together with conditions (A6a) and (A7) above are satisfied for  $X_n$  and  $X$ . The convergence (4.1) therefore holds as required. ■

In complete similarity to the proof of relation (4.1) and to Theorem 3.1, we may deduce, with the help of straightforward extensions of Corollary 10.1, that the following analogues of (4.1) and Theorem 3.1 hold in the present context.

**Theorem 4.2** *Let  $f : \mathbf{R} \rightarrow \mathbf{R}$  be a twice continuously differentiable function such that  $f'$  satisfies the growth condition  $|f'(x)| \leq K(1 + |x|^\alpha)$  for some constants  $K > 0$  and  $\alpha > 0$  and all  $x \in \mathbf{R}$ . Suppose that  $\{(\epsilon_t, \eta_t)\}_{t=0}^\infty$  is a sequence of i.i.d. mean-zero random vectors such that  $E \epsilon_0^2 = \sigma_\epsilon^2$ ,  $E \eta_0^2 = \sigma_\eta^2$ ,  $E \epsilon_0 \eta_0 = \sigma_{\epsilon\eta}$ ,  $E|\epsilon_0|^p < \infty$  and  $E|\eta_0|^p < \infty$  for some with  $p \geq \max(6, 4\alpha)$ . Then*

$$\frac{1}{\sqrt{n}} \sum_{t=2}^{[nr]} f \left( \frac{1}{\sqrt{n}} \sum_{i=1}^{t-1} \epsilon_i \right) \eta_t \rightarrow_d \int_0^r f(W(v)) dV(v). \tag{4.7}$$

Further, using the Phillips-Solo device as in the proof of Theorems 3.1 and 3.1, we obtain the following generalizations of relations (4.1) and (4.7) to the case of linear processes.

**Theorem 4.3** *Suppose that  $w_t = (u_t, v_t)'$  is the linear process  $w_t = G(L)\epsilon_t = \sum_{j=0}^{\infty} G_j \epsilon_{t-j}$ , with  $G(L) = \sum_{j=0}^{\infty} G_j L^j$ ,  $\sum_{j=1}^{\infty} j \|G_j\| < \infty$ ,  $G(1)$  of full rank, and  $\{\epsilon_t\}_{t=0}^{\infty}$  a sequence of i.i.d. mean-zero random vectors such that  $E\epsilon_0 \epsilon_0' = \Sigma_{\epsilon} > 0$  and  $\max_i E|\epsilon_{i0}|^p < \infty$  for some  $p > 4$ . Then*

$$\frac{1}{n} \sum_{t=2}^{[nr]} \left( \sum_{i=1}^{t-1} u_i \right) v_t \rightarrow_d r \lambda_{uv} + \int_0^r W(v) dV(v), \quad (4.8)$$

where  $(W, V) = ((W(s), V(s)), s \geq 0)$  is bivariate Brownian motion with covariance matrix  $\Omega = G(1) \Sigma G(1)$  and  $\lambda_{uv} = \sum_{j=1}^{\infty} E u_0 v_j$ .

Further, if  $f : \mathbf{R} \rightarrow \mathbf{R}$  is a twice continuously differentiable function such that  $f'$  satisfies the growth condition  $|f'(x)| \leq K(1 + |x|^{\alpha})$  for some constants  $K > 0$  and  $\alpha > 0$  and all  $x \in \mathbf{R}$ , and if  $p \geq \max(6, 4\alpha)$ , then

$$\frac{1}{\sqrt{n}} \sum_{t=2}^{[nr]} f \left( \frac{1}{\sqrt{n}} \sum_{i=1}^{t-1} u_i \right) v_t \rightarrow_d \lambda_{uv} \int_0^r f'(W(v)) dv + \int_0^r f(W(v)) dV(v). \quad (4.9)$$

## 5 Asymptotics in stationary and unit root autoregression

This section shows how the martingale convergence approach provides a unified treatment of the limit theory for autoregression as in (5.1) below that includes both stationary ( $\alpha = 0$ ) and unit root ( $\alpha = 1$ ) cases. Let  $(y_t)_{t \in \mathbf{N}}$  be a stochastic process generated in discrete time according to

$$y_t = \alpha y_{t-1} + u_t, \quad (5.1)$$

where  $u_t$  is the linear process  $u_t = C(L)\epsilon_t = \sum_{j=0}^{\infty} c_j \epsilon_{t-j}$ ,  $C(L) = \sum_{j=0}^{\infty} c_j L^j$ ,  $\sum_{j=1}^{\infty} j c_j^2 < \infty$ ,  $C(1) \neq 0$ , and  $(\epsilon_t)_{t \in \mathbf{Z}}$  satisfy assumption (D2) with  $p > 4$ . The initial condition in (5.1) is set at  $t = 0$  and  $y_0$  may be a constant or a random variable. In (5.1) we can use  $\alpha = 0$  to represent the stationary case without loss of generality because  $u_t$  is defined as an arbitrary linear process.

Let  $\hat{\alpha} = \sum_{t=1}^n y_{t-1} y_t / \sum_{t=1}^n y_{t-1}^2$  denote the ordinary least squares (OLS) estimator of  $\alpha$  and let  $t_{\hat{\alpha}}$  be the conventional regression  $t$ -statistic in model (5.1) with  $\alpha = 1$ :  $t_{\hat{\alpha}} = \left( \sum_{t=1}^n y_{t-1}^2 \right)^{1/2} (\hat{\alpha} - 1) / s$ , where  $s^2 = n^{-1} \sum_{t=1}^n (y_t - \hat{\alpha} y_{t-1})^2$ . Further, let  $\hat{\sigma}_u^2$  be a consistent estimator of  $\sigma_u^2 = E u_0^2$  and let  $\hat{\omega}^2$ ,  $\hat{\lambda}$ ,  $\hat{\gamma}$  and  $\hat{\eta}$  be, respectively, consistent nonparametric kernel estimates of the nuisance parameters  $\lambda = \sum_{j=1}^{\infty} E u_0 u_j$ ,  $\omega^2 = \sigma_{\epsilon}^2 C^2(1)$ ,  $\gamma = \sigma_{\epsilon}^2 f_0(1)$  and  $\eta = \left( f_0^2(1) + \sum_{r=1}^{\infty} f_r^2(1) \right)^{1/2}$ , where  $f_0(1) = \sum_{k=0}^{\infty} c_k c_{k+1}$  and  $f_r(1) = \sum_{k=0}^{\infty} c_k c_{k+r-1}$   $r \geq 1$ . Denote by  $Z_{\alpha}$  and  $Z_t$  the statistics  $Z_{\alpha} = n(\hat{\alpha} - 1) - \hat{\lambda} \left( n^{-2} \sum_{t=1}^n y_{t-1}^2 \right)^{-1}$  and  $Z_t = \hat{\sigma}_u \hat{\omega}^{-1} t_{\hat{\alpha}} - \hat{\lambda} \left\{ \hat{\omega} \left( n^{-2} \sum_{t=1}^n y_{t-1}^2 \right)^{1/2} \right\}^{-1}$ .

We prove the following result.

**Theorem 5.1** *If, in model (5.1),  $\alpha = 1$  and  $\sum_{j=1}^{\infty} j |c_j| < \infty$ , then, as  $n \rightarrow \infty$ ,*

$$n(\hat{\alpha} - 1) \rightarrow_d \left( \omega^2 \int_0^1 W(v) dW(v) + \lambda \right) \left( \omega^2 \int_0^1 W^2(v) dv \right)^{-1}, \quad (5.2)$$

$$t_{\hat{\alpha}} \rightarrow_d \sigma_u^{-1} \omega^{-1} \left( \omega^2 \int_0^1 W(v) dW(v) + \lambda \right) \left( \int_0^1 W^2(v) dv \right)^{-1/2}, \quad (5.3)$$

where  $\sigma_u^2 = Eu_0^2$ ,  $\lambda = \sum_{j=1}^{\infty} Eu_0 u_j$  and  $\omega^2 = \sigma_\epsilon^2 C^2(1)$ . One also has the following nuisance-parameter-free limits for the test statistics  $Z_\alpha$  and  $Z_t$  in model (5.1) with  $\alpha = 1$  and  $\sum_{j=1}^{\infty} j|c_j| < \infty$ :

$$Z_\alpha \rightarrow_d \left( \int_0^1 W(v) dW(v) \right) \left( \int_0^1 W^2(v) dv \right)^{-1}, \quad (5.4)$$

$$Z_t \rightarrow_d \left( \int_0^1 W(v) dW(v) \right) \left( \int_0^1 W^2(v) dv \right)^{-1/2}. \quad (5.5)$$

If, in model (5.1),  $\alpha = 0$  and  $\sum_{j=1}^{\infty} j c_j^2 < \infty$ , then, as  $n \rightarrow \infty$ ,

$$\sqrt{n}(\hat{\alpha} - \gamma) \rightarrow_d N(0, \eta^2 / \sigma_u^2), \quad (5.6)$$

$$\frac{\hat{\sigma}_u \sqrt{n}}{\hat{\eta}} (\hat{\alpha} - \gamma) \rightarrow_d N(0, 1). \quad (5.7)$$

**Proof.** Using the continuous mapping theorem (e.g., JS, VI.3.8) and Theorem 2.2 we get  $n^{-2} \sum_{t=1}^n y_{t-1}^2 \rightarrow_d \omega^2 \int_0^1 W^2(v) dv$ , when  $\alpha = 1$ , as in Phillips (1987a). Also, by Theorem 3.1,  $\frac{1}{n} \sum_{t=1}^n y_{t-1} u_t \rightarrow_d \lambda + \omega^2 \int_0^1 W(v) dW(v)$ . These relations then imply by continuous mapping that (5.2) and (5.3) hold. Relations (5.4) and (5.5) are consequences of (5.2) and (5.3). Relations (5.6) and (5.7) follow from Theorem 2.4, the consistency of  $\hat{\eta}$ , and the fact that  $n^{-1} \sum_{t=1}^n u_{t-1}^2 \rightarrow_p \sigma_u^2$  by the law of large numbers. ■

**Remark 5.1** *The martingale convergence approach provides a unifying principle for proving the limit theory in the stationary and unit root cases in the above result. In particular, in the martingale-difference error case (i.e. when Assumption D1 holds and  $u_t = \varepsilon_t$ , allowing for  $\alpha = 1$  or  $|\alpha| < 1$ ) the construction by which the martingale convergence approach is applied is the same in both cases. Thus, in the stationary case we use construction (2.10) and in the unit root case we have a similar construction in (3.7) with  $f(x) = x$  and  $\lambda = 0$ . In the former case, the numerator satisfies a central limit theorem, while in the latter case we have weak convergence to a stochastic integral. This difference makes a unification of the limit theory impossible in terms of existing approaches which rely on central limit arguments in the stationary case and special weak convergence arguments in the unit root case. However, the martingale convergence approach readily accommodates both results and, at the same time, also allows for the difference in the rates of convergence. In effect, in both the stationary and unit root cases, we have convergence of a discrete time martingale to a continuous martingale, thereby unifying the limit theory for autoregression. Section 6 makes this formulation explicit.*

## 6 Unification of the limit theory of autoregression

The present section demonstrates how the martingale convergence approach developed in this paper provides a unified formulation of the limit theory for first order autoregression, including stationary, unit root, local to unity and (together with the conventional martingale convergence theorem) explosive settings.

Specializing (5.1), we consider here the autoregression

$$y_t = \alpha y_{t-1} + \epsilon_t, \quad t = 1, \dots, n \quad (6.1)$$

with martingale-difference errors  $\epsilon_t$  that satisfy assumption (D1) with  $p > 4$ . As in (5.1), the initial condition in (6.1) is set at  $t = 0$  and  $y_0$  may be a constant or a random variable. Extensions to more general initializations are possible but are not considered here to simplify the arguments and notation that follow. We treat the stationary  $|\alpha| < 1$ , unit root  $\alpha = 1$ , local to unity and explosive cases together in what follows and show how the limit theory for all these cases may be formulated in a unified manner within the martingale convergence framework.

We start with the stationary and unit root cases. For  $r \in (0, 1]$ , define the recursive least squares estimator  $\hat{\alpha}_r = \sum_{t=1}^{[nr]} y_{t-1} y_t / \sum_{t=1}^{[nr]} y_{t-1}^2$ , and write

$$\left( \frac{\sum_{t=1}^{[nr]} y_{t-1}^2}{\sigma_\epsilon^2} \right)^{1/2} (\hat{\alpha}_r - \alpha) = \frac{\sum_{t=1}^{[nr]} y_{t-1} \epsilon_t}{\left( \sum_{t=1}^{[nr]} y_{t-1}^2 \sigma_\epsilon^2 \right)^{1/2}} = \frac{X_n(r)}{\left( \tilde{C}'_n(r) \right)^{1/2}}, \quad (6.2)$$

where  $X_n(r)$  is the martingale given by

$$X_n(r) = \begin{cases} \frac{1}{\sqrt{n}} \sum_{t=1}^{[nr]} y_{t-1} \epsilon_t & |\alpha| < 1 \\ \frac{1}{n} \sum_{t=1}^{[nr]} y_{t-1} \epsilon_t & \alpha = 1 \end{cases}, \quad (6.3)$$

and  $\tilde{C}'_n = (\tilde{C}'_n(s), s \geq 0)$  is the modified second characteristic without truncation of  $X_n$  (see JS, Ch. II, §2 and IX.3.25):

$$\tilde{C}'_n(r) = \begin{cases} \frac{1}{n} \sum_{t=1}^{[nr]} y_{t-1}^2 \sigma_\epsilon^2 & |\alpha| < 1 \\ \frac{1}{n^2} \sum_{t=1}^{[nr]} y_{t-1}^2 \sigma_\epsilon^2 & \alpha = 1 \end{cases}. \quad (6.4)$$

By virtue of Remark 2.3 and Theorem 4.1 we have

$$X_n(r) \rightarrow_d X(r) = \begin{cases} \sigma_\alpha \sigma_\epsilon W(r) & |\alpha| < 1 \\ \sigma_\epsilon^2 \int_0^r W(v) dW(v) & \alpha = 1 \end{cases}, \quad (6.5)$$

and

$$\tilde{C}'_n(r) \rightarrow_d C(r) = \begin{cases} \sigma_\alpha^2 \sigma_\epsilon^2 r & |\alpha| < 1 \\ \sigma_\epsilon^4 \int_0^r W(v)^2 dv & \alpha = 1 \end{cases}, \quad (6.6)$$

where  $C = (C(s), s \geq 0)$  is the second predictable characteristic of the continuous martingale  $X$  and  $\sigma_\alpha^2 = 1/(1 - \alpha^2)$ . Thus,

$$\left( \frac{\sum_{t=1}^{[nr]} y_{t-1}^2}{\sigma_\epsilon^2} \right)^{1/2} (\hat{\alpha}_r - \alpha) = \frac{X_n(r)}{\left( \tilde{C}'_n(r) \right)^{1/2}} \rightarrow_d \frac{X(r)}{(C(r))^{1/2}} \quad (6.7)$$

$$= \begin{cases} \frac{1}{\sqrt{1-\alpha^2}} W(r) & |\alpha| < 1 \\ \frac{\int_0^r W(v) dW(v)}{\left( \int_0^r W(v)^2 dv \right)^{1/2}} & \alpha = 1 \end{cases}$$

$$=_d \begin{cases} N(0, 1) & |\alpha| < 1 \\ \frac{\int_0^1 W(v) dW(v)}{(\int_0^1 W(v)^2 dv)^{1/2}} & \alpha = 1 \end{cases},$$

which unifies the limit theory for the stationary and unit root autoregression.

Defining the error variance estimator  $s_r^2 = [nr]^{-1} \sum_{t=1}^{[nr]} (y_t - \hat{\alpha}_r y_{t-1})^2$  and noting that  $s_r^2 \rightarrow_p \sigma_\epsilon^2$  for  $r > 0$ , we have the corresponding limit theory for the recursive  $t$ -statistic

$$t_{\hat{\alpha}}(r) = \left( \frac{\sum_{t=1}^{[nr]} y_{t-1}^2}{s_r^2} \right)^{1/2} (\hat{\alpha}_r - \alpha) = \frac{\sum_{t=1}^{[nr]} y_{t-1} \epsilon_t}{\left( \sum_{t=1}^{[nr]} y_{t-1}^2 \sigma_\epsilon^2 \right)^{1/2}} \frac{\sigma_\epsilon}{s_r} = \frac{X_n(r)}{\left( \tilde{C}'_n(r) \right)^{1/2}} \frac{\sigma_\epsilon}{s_r}$$

$$\rightarrow_d \begin{cases} N(0, 1) & |\alpha| < 1 \\ \frac{\int_0^1 W(v) dW(v)}{(\int_0^1 W(v)^2 dv)^{1/2}} & \alpha = 1 \end{cases}.$$

The theory also extends to cases where  $\alpha$  lies in the neighborhood of unity. In complete similarity to the proof of Theorem 4.1 and to derivations above in this section, one can show that, for  $\alpha = 1 + \frac{c}{n}$ , (6.2) - (6.4) hold with the same normalization as in the unit root case, but in place of (6.5) and (6.6) one now has

$$X_n(r) \rightarrow_d X(r) = \sigma_\epsilon^2 \int_0^r J_c(v) dW(v), \quad \alpha = 1 + \frac{c}{n}, \quad (6.8)$$

$$\tilde{C}'_n(r) \rightarrow_d C(r) = \sigma_\epsilon^4 \int_0^r J_c(v)^2 dv, \quad \alpha = 1 + \frac{c}{n}, \quad (6.9)$$

where  $J_c(v) = \int_0^v e^{c(v-s)} dW(s)$  is a linear diffusion (Phillips, 1987b). We then have

$$\left( \frac{\sum_{t=1}^{[nr]} y_{t-1}^2}{\sigma_\epsilon^2} \right)^{1/2} (\hat{\alpha}_r - \alpha) = \frac{X_n(r)}{\left( \tilde{C}'_n(r) \right)^{1/2}} \rightarrow_d \frac{X(r)}{(C(r))^{1/2}}$$

$$=_d \frac{\int_0^1 J_c(v) dW(v)}{\left( \int_0^1 J_c(v)^2 dv \right)^{1/2}}.$$

Further, when there are moderate deviations from unity of the form  $\alpha = 1 + \frac{c}{n^b}$  for some  $b \in (0, 1)$  and  $c < 0$  (as in Phillips and Magdalinos, 2006, and Giraitis and Phillips, 2006), (6.2) continues to hold but with

$$X_n(r) = \frac{1}{n^{\frac{1+b}{2}}} \sum_{t=1}^{[nr]} y_{t-1} \epsilon_t, \quad \alpha = 1 + \frac{c}{n^b}, \quad c < 0, \quad b \in (0, 1),$$

and  $\tilde{C}'_n(r) = \frac{1}{n^{1+b}} \sum_{t=1}^{[nr]} y_{t-1}^2 \sigma_\epsilon^2$ . Then,  $X_n(r) \rightarrow_d X(r) =_d N\left(0, \frac{\sigma_\epsilon^2}{-2c} r\right)$  and  $\tilde{C}'_n(r) \rightarrow_p C(r) = \frac{\sigma_\epsilon^2}{-2c} r$ . Then, (6.7) again holds with the limit process being  $X(r) / (C(r))^{1/2} =_d N(0, 1)$ .

Next consider the explosive autoregressive case where  $\alpha > 1$ . In this case, (6.2) applies with  $X_n(r) = \frac{1}{\alpha^{[nr]}} \sum_{t=1}^{[nr]} y_{t-1} \epsilon_t$  and  $\tilde{C}'_n(r) = \alpha^{-2[nr]} \sum_{t=1}^{[nr]} y_{t-1}^2 \sigma_\epsilon^2$ . By the martingale convergence theorem,  $\alpha^{-t} y_t \rightarrow_{a.s.}$

$Y_\alpha$ , where  $Y_\alpha = \sum_{s=1}^{\infty} \alpha^{-s} \epsilon_s + y_0$ , and, correspondingly,  $\tilde{C}'_n(r) \rightarrow_{a.s.} C(r) = Y_\alpha^2 \frac{\sigma_\epsilon^2}{\alpha^2 - 1}$ . By further application of the martingale convergence theorem we find that

$$X_n(r) = \frac{1}{\alpha^{[nr]}} \sum_{t=1}^{[nr]} y_{t-1} \epsilon_t = \sum_{t=1}^{[nr]} \frac{y_{t-1}}{\alpha^{t-1}} \frac{\epsilon_t}{\alpha^{[nr]-(t-1)}} \rightarrow_{a.s.} Y_\alpha Z_\alpha, \quad (6.10)$$

with  $Z_\alpha = \sum_{s=1}^{\infty} \alpha^{-s} \epsilon'_s$  where  $(\epsilon'_s)$  is an i.i.d. sequence that is distributionally equivalent to  $(\epsilon_s)$ . In (6.10), the limit of  $X_n(r)$  is the product  $Y_\alpha Z_\alpha$  of the two independent r.v.'s  $Y_\alpha$  and  $Z_\alpha$ . In place of (6.4) we therefore have

$$X_n(r) \rightarrow_{a.s.} X(r) = Y_\alpha Z_\alpha.$$

In place of (6.5) we now have  $\tilde{C}'_n(r) \rightarrow_{a.s.} C(r)$ , where  $C(r)$  denotes  $C(r) = Y_\alpha^2 \sum_{s=1}^{\infty} \alpha^{-2s} \sigma_\epsilon^2 = Y_\alpha^2 \frac{\sigma_\epsilon^2}{\alpha^2 - 1}$ . We therefore find that

$$\begin{aligned} \left( \frac{\sum_{t=1}^{[nr]} y_{t-1}^2}{\sigma_\epsilon^2} \right)^{1/2} (\hat{\alpha}_r - \alpha) &= \frac{X_n(r)}{(\tilde{C}'_n(r))^{1/2}} \rightarrow_{a.s.} \frac{X(r)}{(C(r))^{1/2}} \\ &= \frac{Y_\alpha Z_\alpha}{|Y_\alpha| \left( \frac{\sigma_\epsilon^2}{\alpha^2 - 1} \right)^{1/2}} = \text{sign}(Y_\alpha) \left( \frac{\alpha^2 - 1}{\sigma_\epsilon^2} \right)^{1/2} Z_\alpha. \end{aligned}$$

If  $y_0 = 0$  and  $\epsilon_s$  is i.i.d.  $N(0, \sigma_\epsilon^2)$ , then  $Y_\alpha$  and  $Z_\alpha$  are independent  $N\left(0, \frac{\sigma_\epsilon^2}{\alpha^2 - 1}\right)$  variates and we have

$$\left( \frac{\sum_{t=1}^{[nr]} y_{t-1}^2}{\sigma_\epsilon^2} \right)^{1/2} (\hat{\alpha}_r - \alpha) \rightarrow_{a.s.} \frac{X(r)}{(\tilde{C}'_n(r))^{1/2}} =_d N(0, 1),$$

as shown in early work by White (1958) and Anderson (1959).

## 7 Concluding remarks

The last four sections illustrate the power of the martingale convergence approach in dealing with functional limit theory, weak convergence to stochastic integrals and time series asymptotics for both stationary and nonstationary processes. These examples reveal that the method encompasses much existing asymptotic theory in econometrics and is applicable to a wide class of interesting new problems where the limits involve stochastic integrals and mixed normal distributions. The versatility of the approach is most apparent in the unified treatment that it provides for the limit theory of autoregression, covering stationary, unit root, local to unity and explosive cases. No other approach to the limit theory has yet succeeded in accomplishing this unification.

While the technical apparatus of martingale convergence as it has been developed in Jacod and Shiryaev (2003) is initially somewhat daunting, it should be apparent from these econometric implementations that the machinery has a very broad reach in tackling asymptotic distribution problems in econometrics. Following the example of the applications given here, the methods may be applied directly to deliver asymptotic theory in many interesting econometric models, including models with some roots near unity

and some cointegration as well as models with certain nonlinear forms of cointegration. In addition, the results in the paper can be used in the asymptotic analysis of maximum likelihood estimators for many nonlinear models with integrated time series as well as in the study of weak convergence to stochastic integrals of estimators of various copula parameters, a subject that is receiving increasing attention in the statistical and econometric literature.

## 8 Appendix A1. Background Concepts and Definitions

This appendix briefly reviews some basic notions of semimartingale theory that are used throughout the paper. The processes are defined on a probability space  $(\Omega, \mathfrak{F}, P)$  that is equipped with a filtration  $\mathbb{F} = (\mathfrak{F}_s, s \geq 0)$  of sub- $\sigma$ -fields of  $\mathfrak{F}$ . The definitions formulated below follow the treatment in JS and HWY to make reference to those works more convenient, but they are adapted to the continuous process case that is studied in this paper.

**Definition 8.1** (*Increasing processes*, Definition I.3.1 of JS; Definition III.3.41 of HWY). A real-valued process  $X = (X(s), s \geq 0)$  with  $X(0) = 0$  is called an **increasing** process if all its trajectories are non-negative right-continuous non-decreasing functions.

**Definition 8.2** (*Strong majoration*, Definition VI.3.34 in JS). Let  $X = (X(s), s \geq 0)$  and  $Y = (Y(s), s \geq 0)$  be two real-valued increasing processes. It is said that  $X$  strongly majorizes  $Y$  if the process  $X - Y = (X(s) - Y(s), s \geq 0)$  is itself increasing.

**Definition 8.3** (*Processes with finite variation*, Definition I.3.1 and Proposition I.3.3 in JS; Definition III.3.41 in HWY). A real-valued process  $X = (X(s), s \geq 0)$  is said to be of finite variation if it is the difference of two increasing processes  $Y = (Y(s), s \geq 0)$  and  $Z = (Z(s), s \geq 0)$ , viz.,  $X(s) = Y(s) - Z(s)$ ,  $s \geq 0$ . The process  $\text{Var}(X) = (\text{Var}(X)(s), s \geq 0)$ , where  $\text{Var}(X)(s) = Y(s) + Z(s)$ ,  $s \geq 0$ , is called the variation process of  $X$ .

**Definition 8.4** (*Semimartingales*, Definition I.4.21 in JS; Definition VIII.8.1 in HWY). An  $\mathbf{R}^d$ -valued process  $X = (X(s), s \geq 0)$ ,  $X(s) = (X^1(s), \dots, X^d(s)) \in \mathbf{R}^d$ , is called a  $d$ -dimensional semimartingale with respect to  $\mathbb{F}$  (or a  $d$ -dimensional  $\mathbb{F}$ -semimartingale for short) if, for all  $s \geq 0$  and all  $j = 1, \dots, d$ ,

$$X^j(s) = X^j(0) + M^j(s) + B^j(s), \quad (8.1)$$

where  $X^j(0)$ ,  $j = 1, \dots, d$ , are finite-valued and  $\mathfrak{F}_0$ -measurable r.v.'s,  $M^j = (M^j(s), s \geq 0)$ ,  $j = 1, \dots, d$ , are (real-valued) local  $\mathbb{F}$ -martingales with  $M^j(0) = 0$ ,  $j = 1, \dots, d$ , and  $B^j = (B^j(s), s \geq 0)$ ,  $j = 1, \dots, d$ , are (real-valued)  $\mathbb{F}$ -adapted processes with finite variation.

**Definition 8.5** (*Quadratic variation*, Section I.4e in JS; Section VI.4 in HWY). Let  $M = (M(s), s \geq 0)$  be a continuous square integrable martingale. The **quadratic variation** of  $M$ , denoted  $[M, M]$ , is the unique continuous process  $[M, M] = ([M, M](s), s \geq 0)$ , for which  $M^2 - [M, M]$  is a uniformly integrable martingale which is null at  $s = 0$  (existence and uniqueness of  $[M, M]$  holds by Doob-Meyer decomposition theorem, see Theorem V.5.48 and Section VI.4 in HWY).



Let  $X = (X(s), s \geq 0)$ , where  $X(s) = (X^1(s), \dots, X^d(s)) \in \mathbf{R}^d$ , be a continuous  $d$ -dimensional  $\mathbb{F}$ -semimartingale on  $(\Omega, \mathfrak{F}, P)$ . Then  $X$  admits a unique decomposition (8.1); furthermore, the processes  $B^j = (B^j(s), s \geq 0)$ ,  $j = 1, \dots, d$ , and  $M^j = (M^j(s), s \geq 0)$ ,  $j = 1, \dots, d$ , appearing in (8.1) are continuous (see Lemma I.4.24 in JS).

**Definition 8.6** (*Predictable characteristics of continuous semimartingales, Definition II.2.6 in JS*). The  $\mathbf{R}^d$ -valued process  $B = (B(s), s \geq 0)$ , where  $B(s) = (B^1(s), \dots, B^d(s))$ ,  $s \geq 0$ , is called the **first predictable characteristic** of  $X$ . The  $\mathbf{R}^{d \times d}$ -valued process  $C = (C(s), s \geq 0)$ , where  $C(s) = (C^{ij}(s))_{1 \leq i, j \leq d} \in \mathbf{R}^{d \times d}$ ,  $C^{ij}(s) = [X^i, X^j](s)$ ,  $s \geq 0$ ,  $i, j = 1, \dots, d$ , is called the **second predictable characteristic** of  $X$ .

In the terminology of JS (see Section II.2a in JS),  $X = (X(s), s \geq 0)$  is a semimartingale with the triplet of predictable characteristics  $(B, C, \nu)$ , where the third predictable characteristic of  $X$  (the predictable measure of jumps) is zero in the present case, i.e.,  $\nu = 0$ . Furthermore, since  $X$  is continuous, the triplet does not depend on a truncation function.

The analogues of the concepts in this section for the discrete time case and their versions for general (not necessarily continuous) processes are defined in a similar way (see Hall and Heyde, 1980, JS, Ch. I, §1f, Ch. II, §2 and IX.3.25).

## 9 Appendix A2. Convergence of continuous semimartingales using predictable characteristics

Let  $B = (B(s), s \geq 0)$ ,  $B(s) = (B^1(s), \dots, B^d(s))$ , be an  $\mathbf{R}^d$ -valued process such that  $B^j = (B^j(s), s \geq 0)$ ,  $j = 1, \dots, d$ , are (real-valued)  $\mathbb{F}$ -predictable processes with finite variation and let  $C = (C^{ij})_{1 \leq i, j \leq d}$  be an  $\mathbf{R}^d \times \mathbf{R}^d$ -valued process such that  $C^{ij} = (C^{ij}(s), s \geq 0)$ ,  $i, j = 1, \dots, d$ , are (real-valued)  $\mathbb{F}$ -predictable continuous processes,  $C^{ij}(0) = 0$  and  $C(t) - C(s)$  is a nonnegative symmetric  $d \times d$  matrix for  $s \leq t$ .

**Definition 9.1** (*Martingale problem, Section III.2 in JS*). Let  $X = (X(s), s \geq 0)$ ,  $X(s) = (X^1(s), \dots, X^d(s)) \in \mathbf{R}^d$  be a  $d$ -dimensional continuous process and let  $\mathcal{H}$  denote the  $\sigma$ -field generated by  $X(0)$  and  $\mathcal{L}_0$  denote the distribution of  $X(0)$ . A solution to the **martingale problem** associated with  $(\mathcal{H}, X)$  and  $(\mathcal{L}_0, B, C, \nu)$ , where  $\nu = 0$ , is a probability measure  $P$  on  $(\Omega, \mathfrak{F})$  such that  $X$  is a  $d$ -dimensional  $\mathbb{F}$ -semimartingale on  $(\Omega, \mathfrak{F}, P)$  with the first and second predictable characteristics  $B$  and  $C$ .

Assume that  $(\Omega, \mathfrak{F})$  is the Skorohod space  $(\mathbb{D}(\mathbf{R}_+^d), \mathcal{D}(\mathbf{R}_+^d))$ .

Let  $X_n = (X_n(s), s \geq 0)$ ,  $X_n(s) = (X_n^1(s), \dots, X_n^d(s)) \in \mathbf{R}^d$ ,  $n \geq 1$ , be a sequence of  $d$ -dimensional continuous semimartingales on  $(\Omega, \mathfrak{F}, P)$ . For  $a \geq 0$  and an element  $\alpha = (\alpha(s), s \geq 0)$  of the Skorohod space  $\mathbb{D}(\mathbf{R}_+^d)$ , define, as in IX.3.38 of JS,

$$\begin{aligned} S^a(\alpha) &= \inf\{s : |\alpha(s)| \geq a \text{ or } |\alpha(s-)| \geq a\}, \\ S_n^a &= \inf\{s : |X_n(s)| \geq a\}, \end{aligned} \tag{9.1}$$

where  $\alpha(s-)$  denotes the left-hand limit of  $\alpha$  at  $s$ . For  $r \geq 0$  and  $\alpha \in \mathbb{D}(\mathbf{R}_+^d)$ , denote

$$\bar{\alpha}_{(r)}(x) = \alpha(x - r), \tag{9.2}$$

$x \in \mathbf{R}^d$ .

For  $r \geq 0$  and processes  $B$  and  $C$  introduced at the beginning of the present section, define the processes  $\overline{B}_{(r)} = (\overline{B}_{(r)}(s), s \geq 0)$  and  $\overline{C}_{(r)} = (\overline{C}_{(r)}(s), s \geq 0)$  by

$$\overline{B}_{(r)}(s, \alpha) = B(s + r, \overline{\alpha}_{(r)}) - B(r, \overline{\alpha}_{(r)}), \quad (9.3)$$

$$\overline{C}_{(r)}(s, \alpha) = C(s + r, \overline{\alpha}_{(r)}) - C(r, \overline{\alpha}_{(r)}), \quad (9.4)$$

$\alpha \in \mathbb{D}(\mathbf{R}_+^d), s \geq 0$ .

The following theorem gives sufficient conditions for the weak convergence of a sequence of continuous locally square integrable semimartingales. This theorem, together with Theorem 9.2 below, provides the basis for the study of asymptotic properties of functionals of partial sums in subsequent sections.

Throughout the rest of the section,  $B_n = (B_n(s), s \geq 0)$  and  $C_n = (C_n(s), s \geq 0)$ , where  $B_n(s) = (B_n^1(s), \dots, B_n^d(s))$  and  $C_n(s) = (C_n^{ij}(s))_{1 \leq i, j \leq d}, s \geq 0$ , denote the first and the second predictable characteristics of  $X_n$ , respectively.

In what follows in our initial applications of the martingale convergence argument, both  $X_n$  and  $X$  are continuous. Then, in the corresponding results in JS, the third predictable characteristics of  $X_n$  and  $X$  are zero (i.e.,  $\nu_n = \nu = 0$ ), the first characteristics without truncation of  $X_n$  and  $X$  are the same as  $B_n$  and  $B$  (i.e.,  $B'_n = B_n, B' = B$ ), and the modified characteristics without truncation of  $X_n$  and  $X$  are the same as  $C_n$  and  $C$  (i.e.,  $\tilde{C}'_n = C_n, \tilde{C}' = C$ ). The final section of the paper will consider the case where  $X_n$  has discontinuities and  $X$  is continuous. This extension is particularly valuable in providing a martingale convergence proof of weak convergence of sample covariances to a multivariate stochastic integral.

**Theorem 9.1** (see Theorem IX.3.48, Remark IX.3.40, Theorem III.2.40 and Lemma IX.4.4 in JS and also the proof of Theorem 2.1 in Coffman, Puhalskii and Reiman, 1998). Suppose that the following conditions hold:

(A1) **The local strong majoration hypothesis:** For all  $a \geq 0$ , there is an increasing, deterministic function  $F(a) = (F(s, a), s \geq 0)$  such that the stopped real-valued processes  $(\sum_{j=1}^d \text{Var}(B^j)(s \wedge S^a(\alpha), \alpha), s \geq 0)$  and  $(C^{jj}(s \wedge S_a(\alpha), \alpha), s \geq 0), j = 1, \dots, d$ , are strongly majorized by  $F(a)$  for all  $\alpha \in \mathbb{D}(\mathbf{R}_+^d)$  (see Definitions 8.3 and 8.2).

(A2) **Uniqueness hypothesis:** Let  $\mathcal{H}$  denote the  $\sigma$ -field generated by  $X(0)$  and let  $\mathcal{L}_0$  denote the distribution of  $X(0)$ . For each  $z \in \mathbf{R}^d$  and  $r \geq 0$ , the martingale problem associated with  $(\mathcal{H}, X)$  and  $(\mathcal{L}_0, \overline{B}_{(r)}, \overline{C}_{(r)}, \nu)$ , where  $X(0) = z$  a.s. and  $\nu = 0$ , has a unique solution  $P_{z,r}$  (see Definition 9.1).

(A3) **Measurability hypothesis:** The mapping  $(z, r) \in \mathbf{R}^d \times \mathbf{R}_+ \rightarrow P_{z,r}(A)$  is Borel for all  $A \in \mathfrak{F}$ .

(A4) **The continuity condition:** The mappings  $\alpha \rightarrow B(s, \alpha)$  and  $\alpha \rightarrow C(s, \alpha)$  are continuous for the Skorohod topology on  $\mathbb{D}(\mathbf{R}_+^d)$  for all  $s > 0$ .

(A5)  $X_n(0) \rightarrow_d X(0)$ .

(A6)  $[\mathbf{sup} - \beta_{\text{loc}}] \quad \sup_{0 < s \leq N} |B_n(s \wedge S_n^a) - B(s \wedge S^a(X_n), X_n)| \rightarrow_P 0$  for all  $N \in \mathbf{N}$  and all  $a > 0$ .

$[\gamma_{\text{loc}} - \mathbf{R}_+] \quad C_n(s \wedge S_n^a) - C(s \wedge S^a(X_n), X_n) \rightarrow_P 0$  for all  $s > 0$  and  $a > 0$ .

Then  $X_n \rightarrow_d X$ .

A sufficient condition for (A6) is the following:

$$(A6') \quad [\mathbf{sup} - \beta] \quad \sup_{0 < s \leq N} |B_n(s) - B(s, X_n)| \rightarrow_P 0 \text{ for all } N \in \mathbf{N};$$

$$[\mathbf{sup} - \gamma] \quad \sup_{0 < s \leq N} |C_n(s) - C(s, X_n)| \rightarrow_P 0 \text{ for all } N \in \mathbf{N}.$$

In the case when the limit semimartingale  $X$  satisfies the condition of global strong majoration (see condition (B1) below), conditions (A2)-(A4) and (A6') of Theorem 9.1 simplify and the following result applies.

**Theorem 9.2** (Theorem IX.3.21 in JS). *Suppose that the following conditions hold:*

(B1) **The global strong majoration hypothesis:** *There is an increasing, deterministic function  $F = (F(s), s \geq 0)$  such that the real-valued processes  $(\sum_{j=1}^d \text{Var}(B^j)(s, \alpha), s \geq 0)$  and  $(\sum_{j=1}^d C^{jj}(s, \alpha), s \geq 0)$ ,  $j = 1, \dots, d$ , are strongly majorized by  $F$  for all  $\alpha \in \mathbb{D}(\mathbf{R}_+^d)$  (see Definitions 8.3 and 8.2).*

(B2) **Uniqueness hypothesis:** *Let  $\mathcal{H}$  denote the  $\sigma$ -field generated by  $X(0)$  and let  $\mathcal{L}_0$  denote the distribution of  $X(0)$ . The martingale problem associated with  $(\mathcal{H}, X)$  and  $(\mathcal{L}_0, B, C, \nu)$ , where  $\nu = 0$ , has a unique solution  $P$ .*

(B3) **The continuity condition:** *The mappings  $\alpha \rightarrow B(s, \alpha)$  and  $\alpha \rightarrow C(s, \alpha)$  are continuous for the Skorohod topology on  $\mathbb{D}(\mathbf{R}_+^d)$  for all  $s > 0$ .*

$$(B4) \quad X_n(0) \rightarrow_d X(0).$$

$$(B5) \quad [\mathbf{sup} - \beta] \quad \sup_{0 < s \leq N} |B_n(s) - B(s, X_n)| \rightarrow_P 0 \text{ for all } N \in \mathbf{N};$$

$$[\gamma - \mathbf{R}_+] \quad C_n(s) - C(s, X_n) \rightarrow_P 0 \text{ for all } s > 0.$$

Then  $X_n \rightarrow_d X$ .

The essence of Theorems 9.1 and 9.2 is that convergence of a sequence of semimartingales holds if their predictable characteristics and the initial distributions tend to those of the limit semimartingale (conditions (A5), (A6), (A6'), (B4) and (B5)), the predictable characteristics of the limit process grow in a regular way (conditions (A1) and (B1)) and the process is the only continuous semimartingale with characteristics  $B$  and  $C$  and the given initial distribution (conditions (A2), (A3), (B2)). Technically, conditions (A1), (A5), (A6), (A6') and (B1), (B4) and (B5) guarantee that the sequence  $(X_n)$  is tight and, under conditions (A2)-(A4), (A6), (B2), (B3) and (B5), the limit is identified (see Ch. IX in JS).

One should emphasize here that, whereas the “natural” continuous time “analogue” of the condition on the behavior of variances of partial sums of r.v.’s in the limit theorems in the discrete case might seem to be  $C_n(s) \rightarrow_P C(s, X)$ , in fact one only has to check that  $C_n(s) - C(s, X_n) \rightarrow_P 0$  in Theorems 9.1 and 9.2. The latter condition is simpler because the two components in it are defined on the same probability and only involve  $X_n$  and not the limit process  $X$ .

## 10 Appendix A3. Uniqueness and measurability hypotheses and continuity conditions for homogenous diffusion processes

An important class of limit semimartingales  $X$  for which the conditions of uniqueness and measurability (A2) and (A3) of Theorem 9.1 are satisfied is given by homogenous diffusion processes with infinitesimal

characteristics satisfying quite general conditions. These conditions also assure that the uniqueness hypothesis (B2) of Theorem 9.2 holds. We review some key results from that literature here together with some new results on multivariate diffusion processes that are used in the body of the paper.

For  $d, m \in \mathbf{N}$ , let  $\sigma^{ij} : \mathbf{R}^d \rightarrow \mathbf{R}$ ,  $i = 1, \dots, d$ ,  $j = 1, \dots, m$ , and  $b^i : \mathbf{R}^d \rightarrow \mathbf{R}$ ,  $i = 1, \dots, d$ , be continuous functions and let  $\tilde{W} = (\tilde{W}(s), s \geq 0)$ ,  $\tilde{W}(s) = (W^1(s), \dots, W^m(s))$ , be a standard  $m$ -dimensional Brownian motion. Consider the stochastic differential equation system  $dX^i(s) = \sum_{j=1}^m \sigma^{ij}(X(s))dW^j(s) + b^i(X(s))ds$ ,  $i = 1, \dots, d$ , or, in matrix form,

$$(dX(s))^T = \sigma(X(s))(d\tilde{W}(s))^T + b^T(X(s))ds, \quad (10.1)$$

where  $\sigma : \mathbf{R}^d \rightarrow \mathbf{R}^{d \times m}$  and  $b : \mathbf{R}^d \rightarrow \mathbf{R}^d$  are defined by  $\sigma(x) = (\sigma^{ij}(x))_{1 \leq i \leq d, 1 \leq j \leq m} \in \mathbf{R}^{d \times m}$  and  $b(x) = (b^1(x), \dots, b^d(x)) \in \mathbf{R}^d$ ,  $x \in \mathbf{R}^d$ , and  $y^T$  denotes the transpose of the vector  $y$ .

**Definition 10.1** (see Definition IV.1.2 in Ikeda and Watanabe, 1989, and Definition III.2.24 in JS). A solution to (10.1) is a continuous  $d$ -dimensional process  $X = (X(s), s \geq 0)$ ,  $X(s) = (X^1(s), \dots, X^d(s)) \in \mathbf{R}^d$ , such that, for all  $s \geq 0$  and all  $i = 1, \dots, d$ ,

$$X^i(s) - X^i(0) = \sum_{j=1}^m \int_0^s \sigma^{ij}(X(v))dW^j(v) + \int_0^s b^i(X(v))dv.$$

Such a solution is called a homogenous diffusion process.

**Definition 10.2** (Ikeda and Watanabe, 1989, Definition VI.1.4). It is said that uniqueness of solutions (in the sense of probability laws) holds for (10.1) if, whenever  $X_1$  and  $X_2$  are two solutions for (10.1) such that  $X_1(0) = z$  a.s. and  $X_2(0) = z$  a.s. for some  $z \in \mathbf{R}^d$ , then the laws on the space  $\mathbb{D}(\mathbf{R}_+^d)$  of the processes  $X_1$  and  $X_2$  coincide.

For an element  $\alpha = (\alpha(s), s \geq 0)$  of the Skorohod space  $\mathbb{D}(\mathbf{R}^d)$  and  $i, j = 1, \dots, d$ , define

$$\begin{aligned} B^i(s, \alpha) &= \int_0^s b^i(\alpha(v))dv, \\ C^{ij}(s, \alpha) &= \sum_{k=1}^m \int_0^s \sigma^{ik}(\alpha(v))\sigma^{jk}(\alpha(v))dv = \int_0^s a^{ij}(\alpha(v))dv, \end{aligned} \quad (10.2)$$

where, for  $x \in \mathbf{R}^d$  and  $1 \leq i, j \leq d$ ,

$$a^{ij}(x) = \sum_{k=1}^m \sigma^{ik}(x)\sigma^{jk}(x). \quad (10.3)$$

Further, let  $B(\alpha) = (B(s, \alpha), s \geq 0)$  and  $C(\alpha) = (C(s, \alpha), s \geq 0)$ , where  $B(s, \alpha) = (B^1(s, \alpha), \dots, B^d(s, \alpha))$ , and  $C(s, \alpha) = (C^{ij}(s, \alpha))_{1 \leq i, j \leq d}$ . A solution  $X = (X(s), s \geq 0)$  to equation (10.1) is a semimartingale with the predictable characteristics  $B(X)$  and  $C(X)$ .

The following lemma gives simple sufficient conditions for a homogenous diffusion (a solution to (10.1)) to satisfy continuity conditions (A4) and (B3).

**Lemma 10.1** If  $\sigma(x)$  and  $b(x)$  are continuous in  $x \in \mathbf{R}^d$ , then continuity conditions (A4) and (B3) of Theorems 9.1 and 9.2 are satisfied for the mappings  $\alpha \rightarrow B(s, \alpha)$  and  $\alpha \rightarrow C(s, \alpha)$  defined in (10.2).

**Proof.** The lemma immediately follows from the definition of  $B(s, \alpha)$  and  $C(s, \alpha)$  and continuity of the matrix-valued function  $a(x) = \sigma(x)\sigma^T(x) = (a^{ij}(x))_{1 \leq i, j \leq d}$ , where  $a^{ij}(x)$ ,  $1 \leq i, j \leq d$ , are defined in (10.3). ■

For  $B(s, \alpha)$  and  $C(s, \alpha)$  defined above, one has, in notations (9.3) and (9.4),  $\overline{B}_{(r)}(s, \alpha) = (\overline{B}_{(r)}^1(s, \alpha), \dots, \overline{B}_{(r)}^d(s, \alpha))$  and  $\overline{C}_{(r)}(s, \alpha) = (\overline{C}_{(r)}^1(s, \alpha), \dots, \overline{C}_{(r)}^d(s, \alpha))$ , where

$$\begin{aligned}\overline{B}_{(r)}^i(s, \alpha) &= B^i(s + r, \overline{\alpha}_{(r)}) - B^i(r, \overline{\alpha}_{(r)}) = \\ &= \int_r^{s+r} b^i(\alpha(v - r)) dv = \int_0^s b^i(\alpha(v)) dv = B^i(s, \alpha), \\ \overline{C}_{(r)}^{ij}(s, \alpha) &= C^{ij}(s + r, \overline{\alpha}_{(r)}) - C^{ij}(r, \overline{\alpha}_{(r)}) \\ &= \sum_{k=1}^m \int_r^{s+r} \sigma^{ik}(\alpha(v - r)) \sigma^{jk}(\alpha(v - r)) dv \\ &= \int_0^s \sigma^{ik}(\alpha(u)) \sigma^{jk}(\alpha(v)) dv = C^{ij}(s, \alpha),\end{aligned}\tag{10.4}$$

$i, j = 1, \dots, d$ , that is,  $\overline{B}_{(r)} = B$  and  $\overline{C}_{(r)} = C$  for all  $r \geq 0$  in the uniqueness hypothesis (A2) in Theorem 9.1. Thus, in the case where, in Theorem 9.1, the predictable characteristics of the limit semimartingale  $X$  are  $B(X)$  and  $C(X)$  with  $B$  and  $C$  defined in (10.2) (the limit semimartingale  $X$  is a solution to differential equation (10.1)), conditions (A2) and (A3) simplify to the following:

(A2') **Uniqueness hypothesis:** Let  $\mathcal{H}$  denote the  $\sigma$ -field generated by  $X(0)$  and let  $\mathcal{L}_0$  denote the distribution of  $X(0)$ . For each  $z \in \mathbf{R}^d$ , the martingale problem associated with  $(\mathcal{H}, X)$  and  $(\mathcal{L}_0, B, C, \nu)$ , where  $X(0) = z$  a.s. and  $\nu = 0$ , has a unique solution  $P_z$  (see Definition 9.1).

(A3') **Measurability hypothesis:** The mapping  $z \in \mathbf{R}^d \rightarrow P_z(A)$  is Borel for all  $A \in \mathfrak{S}$ .

The following Theorems 10.1 and 10.2 give sufficient conditions for a homogenous diffusion (a solution to (10.1)) to satisfy conditions (A2) and (A3) (equivalently, (A2') and (A3')). They follow from Theorems IV.2.3, IV.2.4 and IV.3.1 in Ikeda and Watanabe (1989) and Theorem 5.3.1 in Durrett (1996) (see also the discussion following Theorem IV.6.1 on p. 215 in Ikeda and Watanabe, 1989, and Theorem III.2.32 in JS).

**Theorem 10.1** Conditions (A2) and (A3) of Theorem 9.1 are satisfied for a semimartingale  $X = (X(s), s \geq 0)$  with the predictable characteristics  $B(X)$  and  $C(X)$  and  $B$  and  $C$  defined in (10.2) if and only if uniqueness of solutions (in the sense of probability laws) holds for (10.1).

**Theorem 10.2** For any  $z \in \mathbf{R}^d$ , equation (10.1) has a unique (in the sense of probability laws) solution  $X_{(z)} = (X_{(z)}(s), s \geq 0)$  with  $X_{(z)}(0) = z$  if

(C1)  $\sigma(x)$  and  $b(x)$  are locally Lipschitz continuous, that is, for every  $N \in \mathbf{N}$  there exists a constant  $K_N$  such that  $|\sigma(x) - \sigma(y)| + |b(x) - b(y)| \leq K_N|x - y|$  for all  $x, y \in \mathbf{R}^d$  such that  $|x| \leq N$  and  $|y| \leq N$ .

(C2) There is a constant  $K < \infty$  and a function  $\phi(x) \geq 0$ ,  $x \in \mathbf{R}^d$ , with  $\lim_{|x| \rightarrow \infty} \phi(x) = \infty$ , so that if  $X = (X(s), s \geq 0)$  is a solution of (10.1), then  $(e^{-Ks}\phi(X(s)), s \geq 0)$  is a local supermartingale.

Let  $a(x) = \sigma(x)\sigma^T(x)$  (in the component form,  $a(x) = (a^{ij}(x))_{1 \leq i, j \leq d}$ , where  $a^{ij}(x)$  are defined in 10.3). Condition (C2) above holds with  $K = \tilde{K}$  if

(C3)  $\sum_{i=1}^d 2x_i b_i(x) + a_{ii}(x) \leq \tilde{K}(1 + |x|^2)$  for some positive constant  $\tilde{K}$  and all  $x \in \mathbf{R}^d$ .

**Remark 10.1** Analysis of the proof of Theorem 3.1 in Durrett (1996) reveals that condition  $\lim_{|x| \rightarrow \infty} \phi(x) = \infty$  does indeed need to be imposed in the theorem, as indicated in (C2).

**Remark 10.2** Conditions (C1) and (C2) (and, thus, (C1) and (C3)) of Theorem 10.2 guarantee existence of a global solution to (10.1) (that is, a solution defined for all  $s \in \mathbf{R}_+$ ) and its uniqueness. Formally, for any  $x \in \mathbf{R}$ , a solution  $X_{(x)}$  to (10.1) with the initial condition  $X_{(x)}(0) = x$  and the stopping times  $\tilde{S}_n$  defined by  $\tilde{S}_n = \inf\{s \geq 0 : |X_{(x)}(s)| \geq n\}$ , one has that the explosion time  $\tilde{S}$  for  $X_{(x)}$  given by  $\tilde{S} = \lim_{n \rightarrow \infty} \tilde{S}_n$  is infinite a.s.:  $\tilde{S} = \infty$  a.s.

**Remark 10.3** In fact, conditions (C1) and (C2) (and, thus, (C1) and (C3)) of Theorem 10.2 are sufficient not only for existence and uniqueness of solutions for (10.1) in the sense of probability laws (Definition 10.2), but also for pathwise uniqueness of solutions (see Ikeda and Watanabe, 1989, Ch. IV). Theorems 10.1 and 10.2 have a counterpart, due to Stroock and Varadhan, according to which existence and uniqueness of solutions in the sense of probability laws holds for (10.1) if the following conditions are satisfied:

(C1')  $b(x)$  is bounded;

(C2')  $a(x) = \sigma(x)\sigma^T(x)$  is bounded and continuous and everywhere invertible.

(see Theorem IV.3.3 and the discussion following Theorem IV.6.1 on p. 215 in Ikeda and Watanabe, 1989, Theorem III.2.34 and Corollary III.2.41 in JS, and Chapters 6 and 7 in Stroock and Varadhan, 1979).

For the proof of the main results in the paper, we will need a corollary of Theorems 10.1 and 10.2 in the case  $d = 2$  and  $m = 1$  (that is, in the case of a two-dimensional homogenous diffusion driven by a single Brownian motion) and functions  $\sigma : \mathbf{R}^2 \rightarrow \mathbf{R}^{2 \times 1}$  and  $b : \mathbf{R}^2 \rightarrow \mathbf{R}^2$  given by

$$\begin{aligned}\sigma(x_1, x_2) &= (g_1(x_2), 1)^T, \\ b(x_1, x_2) &= (g_2(x_2), 0),\end{aligned}\tag{10.5}$$

where  $g_i : \mathbf{R} \rightarrow \mathbf{R}$ ,  $i = 1, 2$ , are some continuous functions. In other words, we consider the stochastic differential equation

$$\begin{aligned}dX_1(s) &= g_1(X_2(s))dW(s) + g_2(X_2(s))ds; \\ dX_2(s) &= dW(s).\end{aligned}\tag{10.6}$$

A solution  $X = (X(s), s \geq 0)$ ,  $X(s) = (X_1(s), X_2(s))$  to (10.6) is a two-dimensional semimartingale with the predictable characteristics  $B(X)$  and  $C(X)$ , where, for an element  $\alpha = (\alpha(s), s \geq 0)$ ,  $\alpha(s) =$

$(\alpha_1(s), \alpha_2(s))$  of the Skorohod space  $\mathbb{D}(\mathbf{R}_+^2)$ ,

$$\begin{aligned} B(s, \alpha) &= \left( \int_0^s g_2(\alpha_2(v)) dv, 0 \right) = (B^1(s, \alpha), B^2(s, \alpha)), \\ C(s, \alpha) &= \begin{pmatrix} \int_0^s g_1^2(\alpha_2(v)) dv & \int_0^s g_1(\alpha_2(v)) dv \\ \int_0^s g_1(\alpha_2(v)) dv & s \end{pmatrix} = \begin{pmatrix} C^{11}(s, \alpha) & C^{12}(s, \alpha) \\ C^{21}(s, \alpha) & C^{22}(s, \alpha) \end{pmatrix}. \end{aligned} \quad (10.7)$$

**Corollary 10.1** *Suppose that the conditions hold:*

( $\tilde{C}1$ ) *The functions  $g_1$  and  $g_2$  are locally Lipschitz continuous, that is, for every  $N \in \mathbf{N}$  there exists a constant  $K_N$  such that  $|g_i(x) - g_i(y)| \leq K_N|x - y|$ ,  $i = 1, 2$ , for all  $x, y \in \mathbf{R}$  such that  $|x| \leq N$  and  $|y| \leq N$ ;*

( $\tilde{C}2$ )  *$g_1$  and  $g_2$  satisfy the growth condition*

$$|g_i(x)| \leq e^{K|x|}, \quad i = 1, 2, \quad (10.8)$$

*for some positive constant  $K$  and all  $x \in \mathbf{R}$ .*

*Then, for any  $z \in \mathbf{R}^2$ , stochastic differential equation (10.6) has a unique solution  $X_{(z)} = (X_{(z)}(s), s \geq 0)$  with  $X_{(z)}(0) = z$  and, thus, by Theorem 10.1, conditions (A2) and (A3) of Theorem 9.1 are satisfied for a semimartingale  $X = (X(s), s \geq 0)$ ,  $X(s) = (X_1(s), X_2(s))$  with the predictable characteristics  $B(X)$  and  $C(X)$  and  $B$  and  $C$  defined in (10.7).*

**Proof.** Clearly, under the assumptions of the corollary, condition (C1) of Theorem 10.2 is satisfied for the mappings  $\sigma$  and  $b$  defined in (10.5). Let us show that condition (C2) of Theorem 10.2 is satisfied with  $A = 2 + 2K^2$  and  $\phi(x_1, x_2) = x_1^2 + e^{2Kx_2} + e^{-2Kx_2}$ . Clearly,  $\lim_{|(x_1, x_2)| \rightarrow \infty} \phi(x_1, x_2) = \infty$ . Similar to the proof of Theorem 5.3.1 in Durrett (1996), by Itô's formula we have that

$$\begin{aligned} d \left[ e^{-As} \phi(X_1(s), X_2(s)) \right] &= e^{-As} \left[ -A \left( X_1^2(s) + e^{2KX_2(s)} + e^{-2KX_2(s)} \right) \right. \\ &\quad \left. + 2X_1(s)g_2(X_2(s)) + g_1^2(X_2(s)) + 2K^2 \left( e^{2KX_2(s)} + e^{-2KX_2(s)} \right) \right] ds \\ &\quad + e^{-As} \left[ 2X_1(s)g_1(X_2(s)) + 2K \left( e^{2KX_2(s)} - e^{-2KX_2(s)} \right) \right] dW(s). \end{aligned}$$

Since

$$\begin{aligned} &-A \left( X_1^2(s) + e^{2KX_2(s)} + e^{-2KX_2(s)} \right) + 2X_1(s)g_2(X_2(s)) + g_1^2(X_2(s)) + \\ &2K^2 \left( e^{2KX_2(s)} + e^{-2KX_2(s)} \right) = -AX_1^2(s) + 2X_1(s)g_2(X_2(s)) + g_1^2(X_2(s)) - \\ &2 \left( e^{2KX_2(s)} + e^{-2KX_2(s)} \right) \leq (1 - A)X_1^2(s) + g_2^2(X_2(s)) + g_1^2(X_2(s)) - \\ &2 \left( e^{2KX_2(s)} + e^{-2KX_2(s)} \right) \leq 0 \end{aligned}$$

by condition ( $\tilde{C}2$ ), we have that the process  $(e^{-s}\phi(X(s)), s \geq 0)$  is a local supermartingale. Consequently, (C2) indeed holds and, by Theorems 10.1 and 10.2, the proof is complete. ■

**Remark 10.4** *It is important to note that condition (C2') of Remark 10.3 is not satisfied for stochastic differential equation (10.6) since, as it is easy to see, the matrix  $a(x) = \sigma(x)\sigma^T(x)$  is degenerate for  $\sigma$  defined in (10.5). The same applies, in general, to condition (C3) of Theorem 10.2. Therefore, the counterpart to Theorems 10.1 and 10.2 given by Remark 10.3 and, in general, linear growth condition (C3) cannot be employed to justify uniqueness and measurability hypothesis of Theorem 9.1 for the limit martingale  $X$  with the predictable characteristics  $B(X)$  and  $C(X)$  and  $B$  and  $C$  defined in (10.7). This is crucial in the proof of convergence to stochastic integrals in Section 3, where the limit semimartingales are solutions to (10.6), and we employ the result given by Corollary 10.1 to justify that conditions (A2) and (A3) of Theorem 9.1 hold for them.*

The following is a straightforward corollary of Lemma 10.1 in the case of stochastic equation (10.6).

**Corollary 10.2** *Continuity conditions (A4) and (B3) of Theorems 9.1 and 9.2 hold for the mappings  $\alpha \rightarrow B(s, \alpha)$  and  $\alpha \rightarrow C(s, \alpha)$  defined in (10.7) if the functions  $g_1(x)$  and  $g_2(x)$  are continuous (in particular, (A4) and (B3) hold under assumption of local Lipschitz continuity (C1) of Corollary 10.1).*

## 11 Appendix A4. Embedding of a martingale into a Brownian motion

The following lemma gives the Skorohod embedding of martingales and a strong approximation to their quadratic variation. It was obtained in Park and Phillips (1999) in the case of the space  $\mathbb{D}([0, 1])$  (see also Theorem A.1 in Hall and Heyde, 1980, Phillips and Ploberger, 1996, and Park and Phillips, 2001). The argument in the case of the space  $\mathbb{D}(\mathbf{R}_+)$  is the same as in Park and Phillips (1999).

**Lemma 11.1** *(Park and Phillips, 1999, Lemma 6.2). Let assumption (D1) hold.<sup>5</sup> Then there exists a probability space supporting a standard Brownian motion  $W$  and an increasing sequence of nonnegative stopping times  $(T_k)_{k \geq 0}$  with  $T_0 = 0$  such that*

$$\frac{1}{\sqrt{n}} \sum_{k=1}^t \epsilon_k =_d W\left(\frac{T_t}{n}\right), \quad (11.1)$$

$t \in \mathbf{N}$ , and

$$\max_{1 \leq t \leq Nn} \frac{|T_t - \sigma_\epsilon^2 t|}{n^q} \rightarrow_{a.s.} 0, \quad (11.2)$$

$$\sup_{0 \leq r \leq N} \left| \frac{T_{[nr]}}{n} - \sigma_\epsilon^2 r \right| =_{a.s.} o(n^{q-1}) \quad (11.3)$$

for all  $N \in \mathbf{N}$  and any  $q > \max(1/2, 2/p)$ . In addition to the above,  $T_t$  is  $\mathcal{E}_t$ -measurable and, for all  $\beta \in [1, p/2]$ ,

$$E((T_t - T_{t-1})^\beta | \mathcal{E}_{t-1}) \leq K_\beta E(|\epsilon_t|^{2\beta} | \mathfrak{F}_{t-1}) \quad a.s.$$

for some constant  $K_\beta$  depending only on  $\beta$ ,

$$E(T_t - T_{t-1} | \mathcal{E}_{t-1}) = \sigma_\epsilon^2 \quad a.s.,$$

where  $\mathcal{E}_t$  is the  $\sigma$ -field generated by  $(\epsilon_k)_{k=1}^t$  and  $W(s)$  for  $0 \leq s \leq T_t$ .

---

<sup>5</sup>As in assumption (D1), below,  $(\mathfrak{F}_t)$  denotes a natural filtration for  $(\epsilon_t)$ .



## 12 Appendix A5. Auxiliary lemmas

**Lemma 12.1** (*Billingsley, 1968, Theorem 4.1*). Let  $(\Omega, \mathfrak{F}, P)$  be a probability space and let  $(E, \mathcal{E})$  be a metric space with a metric  $\rho$ . Let  $X_n, Y_n, n \geq 1$ , and  $X$  be  $E$ -valued random elements on  $(\Omega, \mathfrak{F}, P)$  such that  $X_n \rightarrow_d X$  and  $\rho(X_n, Y_n) \rightarrow_P 0$ . Then  $Y_n \rightarrow_d X$ .

For  $\alpha, \beta \in \mathbb{D}(\mathbf{R}_+)$  such that  $\beta(s) \geq 0$  for  $s \in \mathbf{R}_+$  let  $\alpha \circ \beta \in \mathbb{D}(\mathbf{R}_+)$  denote the composition of  $\alpha$  and  $\beta$ , that is, the function  $(\alpha \circ \beta)(s) = \alpha(\beta(s))$ ,  $s \geq 0$ .

**Lemma 12.2** Suppose that  $X_n \rightarrow_d X$  and  $Y_n \rightarrow_P Y$ , where  $X = (X(s), s \geq 0)$  and  $Y = (Y(s), s \geq 0)$  are continuous processes and  $X(s) \geq 0$  for  $s \in \mathbf{R}_+$ . Then  $X_n \circ Y_n \rightarrow_d X \circ Y$ .

For the proof of Lemma 12.2, we need the following well-known result. Let  $\rho(x, y)$  denote the Skorohod metric on  $\mathbb{D}(\mathbf{R}_+)$  and let  $\mathbb{C}(\mathbf{R}_+)$  denote the space of continuous functions on  $\mathbf{R}_+$ .

**Lemma 12.3** (*Proposition VI.1.17 in JS; see also Theorem 15.12 in HWY*). Let  $x_n \in \mathbb{D}(\mathbf{R}_+)$ ,  $n \geq 1$ , and  $x \in \mathbb{D}(\mathbf{R}_+)$ . Then

$$\sup_{0 \leq s \leq N} |x_n(s) - x(s)| \rightarrow 0 \quad (12.1)$$

for all  $N \in \mathbf{N}$  implies that

$$\rho(x_n, x) \rightarrow 0. \quad (12.2)$$

If, in addition,  $x \in \mathbb{C}(\mathbf{R}_+)$ , then relations (12.1) and (12.2) are equivalent.

**Proof of Lemma 12.2.** Relations  $X_n \rightarrow_d X$  and  $Y_n \rightarrow_P Y$  imply (see Theorem 4.4 in Billingsley, 1968) that

$$(X_n, Y_n) \rightarrow_d (X, Y). \quad (12.3)$$

It is not difficult to see that the mapping  $\psi : \mathbb{D}(\mathbf{R}_+^2) \rightarrow \mathbb{D}(\mathbf{R}_+)$  defined by  $\psi(\alpha, \beta) = \alpha \circ \beta$  for  $(\alpha, \beta) \in \mathbb{D}(\mathbf{R}_+^2)$  with  $\beta(s) \geq 0$ ,  $s \in \mathbf{R}_+$ , is continuous at  $(\alpha, \beta)$  such that  $\alpha, \beta \in \mathbb{C}(\mathbf{R}_+)$ . Indeed suppose that, for the Skorohod metric  $\rho$ ,  $\rho(\alpha_n, \alpha) \rightarrow 0$  and  $\rho(\beta_n, \beta) \rightarrow 0$ , where  $\alpha_n, \beta_n \in \mathbb{D}(\mathbf{R}_+)$ ,  $n \geq 1$ , and  $\alpha, \beta \in \mathbb{C}(\mathbf{R}_+)$ . We have that, for any  $N \in \mathbf{N}$ ,

$$\begin{aligned} \sup_{0 \leq s \leq N} |\alpha_n \circ \beta_n(s) - \alpha \circ \beta(s)| &\leq \sup_{0 \leq s \leq N} |\alpha_n \circ \beta_n(s) - \alpha \circ \beta_n(s)| + \\ &\sup_{0 \leq s \leq N} |\alpha \circ \beta_n(s) - \alpha \circ \beta(s)| \end{aligned} \quad (12.4)$$

Using Lemma 12.3 with  $x_n = \beta_n$  and  $x = \beta$  and continuity of  $\beta$  we get that, for all  $n \geq 1$ ,  $\sup_{0 \leq s \leq N} |\beta_n(s)| \leq \sup_{0 \leq s \leq N} |\beta_n(s) - \beta(s)| + \sup_{0 \leq s \leq N} |\beta(s)| \leq K(N) < \infty$ . Consequently, from the same lemma with  $x_n = \alpha_n$  and  $x = \alpha$  it follows that, for all  $N \in \mathbf{N}$ ,

$$\sup_{0 \leq s \leq N} |\alpha_n \circ \beta_n(s) - \alpha \circ \beta_n(s)| \leq \sup_{0 \leq s \leq K(N)} |\alpha_n(s) - \alpha(s)| \rightarrow 0. \quad (12.5)$$

Using again Lemma 12.3 with  $x_n = \beta_n$  and  $x = \beta$  and uniform continuity of  $\alpha$  on compacts we also get that, for all  $N \in \mathbf{N}$ ,

$$\sup_{0 \leq s \leq N} |\alpha \circ \beta_n(s) - \alpha \circ \beta(s)| \rightarrow 0. \quad (12.6)$$

Relations (12.4)-(12.6) imply that (12.1) holds with  $x_n = \alpha_n \circ \beta_n$  and  $x = \alpha \circ \beta$  and thus, by Lemma 12.3,  $\rho(\alpha_n \circ \beta_n, \alpha \circ \beta) \rightarrow 0$ , as required.

Continuity of  $\psi$  and property (12.3) imply, by continuous mapping theorem (see JS, VI.3.8, and Billingsley, 1968, Corollary 1 to Theorem 5.1 and the discussion on pp. 144-145) that  $X_n \circ Y_n = \psi(X_n, Y_n) \rightarrow_d \psi(X, Y) = X \circ Y$ . ■

**Lemma 12.4** *Let  $p > 0$ . Suppose that a sequence of identically distributed r.v.'s  $(\xi_t)_{t \in \mathbf{N}_0}$  is such that  $E|\xi_0|^p < \infty$ . Then*

$$n^{-1/p} \max_{0 \leq k \leq nN} |\xi_k| \rightarrow_P 0 \quad (12.7)$$

for all  $N \in \mathbf{N}$ .

**Proof.** Evidently, (12.7) is equivalent to  $n^{-1} \max_{0 \leq k \leq nN} |\xi_k|^p \rightarrow_P 0$ . Similar to the discussion preceding Theorem 3.4 in Phillips and Solo (1992) and the discussion in Hall and Heyde (1980, p. 53) we get that this relation, in turn, is equivalent to

$$J_n = \frac{1}{n} \sum_{k=1}^{Nn} |\xi_k|^p I(|\xi_k|^p > n\delta) \rightarrow_P 0$$

for all  $\delta > 0$ . The latter property holds because  $EJ_n \leq NE|\xi_0|^p I(|\xi_0|^p > n\delta) \rightarrow 0$  by the dominated convergence theorem (see Theorem A.7 in Hall and Heyde, 1980) since  $E|\xi_0|^p < \infty$ . ■

As it is well known, the conclusion of Lemma 12.7 can be strengthened in the case of martingales. In particular, the following lemma holds.

**Lemma 12.5** *Suppose that  $(\eta_{tn}, \mathfrak{F}_t)_{t \in \mathbf{N}}$ ,  $n \geq 1$ , is an array of martingale-difference sequences with  $\max_{1 \leq t \leq nN} E\eta_{tn}^2 \leq L$  for some constant  $L > 0$  and all  $n, N \in \mathbf{N}$ . Then*

$$n^{-1} \max_{1 \leq k \leq nN} \left| \sum_{t=1}^k \eta_{tn} \right| \rightarrow_P 0$$

for all  $N \in \mathbf{N}$ .

**Proof.** By Kolmogorov's inequality for martingales (Hall and Heyde, 1980, Corollary 2.1) we get that, for all  $\delta > 0$ ,

$$P\left(n^{-1} \max_{1 \leq k \leq nN} \left| \sum_{t=1}^k \eta_{tn} \right| > \delta\right) \leq E\left(\sum_{t=1}^{Nn} \eta_{tn}\right)^2 / (\delta^2 n^2) \leq$$

$$N \max_{1 \leq t \leq Nn} E\eta_{tn}^2 / n \leq NL/n \rightarrow 0,$$

as required. ■

**Lemma 12.6** For the r.v.'s  $\tilde{\epsilon}_t$  defined in the proof of Theorem 2.2, one has  $E|\tilde{\epsilon}_0|^p < \infty$  if  $(\epsilon_t)_{t \in \mathbf{Z}}$  satisfy assumption (D2) with  $p > 2$ .

**Proof.** Since  $E|\epsilon_0|^p < \infty$ , by the triangle inequality for the  $L_p$ -norm  $\|\cdot\|_p = (E|\cdot|^p)^{1/p}$  and Lemma 2.1 in Phillips and Solo (1992) we have  $\|\tilde{\epsilon}_0\|_p = \|\sum_{j=0}^{\infty} \tilde{\epsilon}_j \epsilon_{-j}\|_p \leq \|\epsilon_0\|_p \sum_{j=0}^{\infty} |\tilde{\epsilon}_j| < \infty$ . ■

**Lemma 12.7** For  $g_{jk}$  defined in the proof of Theorem 2.4, one has  $\sum_{k=0}^{\infty} \sum_{j=k+1}^{\infty} |g_{rj}| < \infty$  for all  $r$  if  $\sum_{j=1}^{\infty} j c_j^2 < \infty$ .

**Proof.** Using change of summation indices and Hölder inequality, we have that

$$\begin{aligned} \sum_{k=0}^{\infty} \sum_{j=k+1}^{\infty} |g_{rj}| &= \sum_{k=0}^{\infty} \sum_{j=k+1}^{\infty} |c_j| |c_{j+r}| = \sum_{j=1}^{\infty} j |c_j| |c_{j+r}| = \\ &= \sum_{j=1}^{\infty} j^{1/2} |c_j| j^{1/2} |c_{j+r}| \leq \left( \sum_{j=1}^{\infty} j |c_j|^2 \right)^{1/2} \left( \sum_{j=1}^{\infty} j |c_{j+r}|^2 \right)^{1/2} < \infty, \end{aligned}$$

as required. ■

**Lemma 12.8** For the r.v.'s  $\tilde{u}_{at}$  and  $\tilde{u}_{bt}$  defined in the proof of Theorem 2.4, one has  $Eu_{a0}^2 < \infty$  and  $Eu_{b0}^2 < \infty$  if  $(\epsilon_t)_{t \in \mathbf{Z}}$  satisfy assumption (D2) with  $p > 2$ .

**Proof.** The property  $Eu_{b0}^2 < \infty$  holds by Lemma 5.9 in Phillips and Solo (1992). By the triangle inequality for the  $L_2$ -norm  $\|\cdot\|_2 = (E(\cdot)^2)^{1/2}$  and Lemma 12.7,  $\|\tilde{u}_{a0}\|_2 = \left\| \sum_{k=0}^{\infty} \tilde{g}_{mk} \epsilon_{-k}^2 \right\|_2 \leq \|\epsilon_0^2\|_2 \sum_{k=0}^{\infty} \sum_{j=k+1}^{\infty} |g_{mj}| < \infty$  Consequently,  $E\tilde{u}_{a0}^2 = O\left(\sum_{k=0}^{\infty} \sum_{j=k+1}^{\infty} |g_{mj}|\right)^2 < \infty$ . ■

**Lemma 12.9** For  $\tilde{h}_{kr}$  defined in the proof of Theorem 3.1, one has  $\sum_{r=0}^{\infty} \sum_{k=0}^{\infty} |\tilde{h}_{kr}| < \infty$  if  $\sum_{j=1}^{\infty} j |c_j| < \infty$ .

**Proof.** By definition of  $\tilde{h}_{kr}$ , it suffices to prove that

$$\sum_{r=0}^{\infty} \sum_{k=0}^{\infty} \sum_{j=k+1}^{\infty} |c_j| |\tilde{\epsilon}_{j+r}| < \infty \quad (12.8)$$

and

$$\sum_{r=0}^{\infty} \sum_{k=0}^{\infty} \sum_{j=k+1}^{\infty} |\tilde{\epsilon}_j| |c_{j+r}| < \infty. \quad (12.9)$$

Using change of summation indices, we have that

$$\sum_{r=0}^{\infty} \sum_{k=0}^{\infty} \sum_{j=k+1}^{\infty} |c_j| |\tilde{\epsilon}_{j+r}| \leq \sum_{r=0}^{\infty} \sum_{j=1}^{\infty} j |c_j| |\tilde{\epsilon}_{j+r}| =$$

$$\sum_{j=1}^{\infty} j|c_j| \sum_{k=j}^{\infty} |\tilde{c}_k| \leq \left( \sum_{j=1}^{\infty} j|c_j| \right) \left( \sum_{k=1}^{\infty} |\tilde{c}_k| \right) < \infty, \quad (12.10)$$

$$\sum_{r=0}^{\infty} \sum_{k=0}^{\infty} \sum_{j=k+1}^{\infty} |\tilde{c}_j| |c_{j+r}| \leq \sum_{r=0}^{\infty} \sum_{j=1}^{\infty} j |\tilde{c}_j| |c_{j+r}| \leq \sum_{r=0}^{\infty} \sum_{j=1}^{\infty} j |c_{j+r}| \sum_{k=j+1}^{\infty} |c_k| \leq$$

$$\sum_{r=0}^{\infty} \sum_{j=1}^{\infty} |c_{j+r}| \sum_{k=j+1}^{\infty} k |c_k| \leq \left( \sum_{j=1}^{\infty} \sum_{s=j}^{\infty} |c_s| \right) \left( \sum_{k=1}^{\infty} k |c_k| \right) < \infty \quad (12.11)$$

because, as in Lemma 2.1 in Phillips and Solo (1992) and its proof,  $\sum_{j=1}^{\infty} j|c_j| < \infty$  implies that  $\sum_{j=1}^{\infty} |\tilde{c}_j| < \infty$  and, even stronger,  $\sum_{j=1}^{\infty} \sum_{s=j}^{\infty} |c_s| < \infty$ . ■

**Lemma 12.10** *For the r.v.'s  $\tilde{w}_{ak}$  and  $\tilde{w}_{bk}$  defined in the proof of Theorem 3.1, one has  $E|\tilde{w}_{a0}|^{p/2} < \infty$  and  $E|\tilde{w}_{b0}|^{p/2} < \infty$  if  $(\epsilon_t)_{t \in \mathbf{Z}}$  satisfy assumption (D2) with  $p > 2$  and  $\sum_{j=1}^{\infty} j|c_j| < \infty$ .*

**Proof.** Denote  $q = p/2$ . Since  $E|\epsilon_0|^p < \infty$ , by the triangle inequality for the  $L_q$ -norm  $\|\cdot\|_q = (E|\cdot|^q)^{1/q}$  and Lemma 12.9, we get

$$\|\tilde{w}_{a0}\|_q = \left\| \sum_{k=0}^{\infty} \tilde{h}_{k0} \epsilon_{-k}^2 \right\|_q \leq \|\epsilon_0\|_p \sum_{k=0}^{\infty} |\tilde{h}_{k0}| < \infty,$$

$$\|\tilde{w}_{b0}\|_q \leq \sum_{r=1}^{\infty} \|\tilde{h}_r(L) \epsilon_0 \epsilon_{-r}\|_q \leq (\|\epsilon_0\|_q)^2 \sum_{r=1}^{\infty} \sum_{k=0}^{\infty} |\tilde{h}_{kr}| < \infty.$$

Consequently,  $E|\tilde{w}_{a0}|^q < \infty$  and  $E|\tilde{w}_{b0}|^q < \infty$ , as required. ■

**Lemma 12.11** *For the r.v.'s  $\eta_{t-1}^h$  defined in the proof of Theorem 3.1, one has  $E(\eta_{-1})^4 < \infty$  if  $(\epsilon_t)_{t \in \mathbf{Z}}$  satisfy assumption (D2) with  $p \geq 4$  and  $\sum_{j=1}^{\infty} j|c_j| < \infty$ .*

**Proof.** As in Lemma 2.1 in Phillips and Solo (1992) and its proof,  $\sum_{j=1}^{\infty} j|c_j| < \infty$  implies that  $\sum_{j=1}^{\infty} |\tilde{c}_j| < \infty$  and, even stronger,  $\sum_{j=1}^{\infty} \sum_{s=j}^{\infty} |c_s| < \infty$ . Therefore, under the assumptions of the theorem,

$$\sum_{r=1}^{\infty} |h_r(1)| \leq \sum_{r=1}^{\infty} \sum_{k=0}^{\infty} |c_k| |\tilde{c}_{k+r}| + \sum_{r=1}^{\infty} \sum_{k=0}^{\infty} |\tilde{c}_k| |c_{k+r}| \leq 2 \left( \sum_{j=0}^{\infty} |c_j| \right) \left( \sum_{j=0}^{\infty} |\tilde{c}_j| \right) < \infty.$$

Using the triangle inequality for the  $L_4$ -norm  $\|\cdot\|_4 = (E|\cdot|^4)^{1/4}$ , we get, therefore,

$$\|\eta_{-1}\|_4 = \left\| \sum_{r=1}^{\infty} h_r(1) \epsilon_{-r} \right\|_4 \leq \|\epsilon_0\|_4 \sum_{r=1}^{\infty} |h_r(1)| < \infty.$$

Consequently,  $E(\epsilon_{-1}^h)^4 = O(\sum_{r=1}^{\infty} h_r(1)) < \infty$ . ■

**Lemma 12.12** *Under the assumptions of Theorem 3.1 one has*

$$\max_{1 \leq k \leq nN} E \left( f' \left( \frac{1}{\sqrt{n}} \sum_{t=1}^k u_t \right) \right)^4 \leq L,$$

for some constant  $L > 0$  and all  $n, N \in \mathbf{N}$ .

**Proof.** The growth condition  $|f'(x)| \leq K(1 + |x|^\alpha)$  evidently implies that  $(f'(x))^4 \leq K(1 + x^{4\alpha})$ . Consequently, using (2.6), we get that, for all  $k$ ,

$$\begin{aligned} \left( f' \left( \frac{1}{\sqrt{n}} \sum_{t=1}^k u_t \right) \right)^4 &\leq K \left( 1 + \left| \frac{1}{\sqrt{n}} \sum_{t=1}^k u_t \right|^{4\alpha} \right) = \\ &K \left( 1 + \left| \frac{C(1)}{\sqrt{n}} \sum_{t=1}^k \epsilon_t + \frac{\tilde{\epsilon}_0}{\sqrt{n}} - \frac{\tilde{\epsilon}_k}{\sqrt{n}} \right|^{4\alpha} \right) \leq \\ &K \left( 1 + \left| \frac{C(1)}{\sqrt{n}} \sum_{t=1}^k \epsilon_t \right|^{4\alpha} + \left| \frac{\tilde{\epsilon}_0}{\sqrt{n}} \right|^{4\alpha} + \left| \frac{\tilde{\epsilon}_k}{\sqrt{n}} \right|^{4\alpha} \right). \end{aligned}$$

Thus, for some constant  $K > 0$ ,

$$\begin{aligned} \max_{1 \leq k \leq nN} E \left( f' \left( \frac{1}{\sqrt{n}} \sum_{t=1}^k u_t \right) \right)^4 &\leq \\ K \left( 1 + \max_{1 \leq k \leq nN} E \left| \frac{C(1)}{\sqrt{n}} \sum_{t=1}^k \epsilon_t \right|^{4\alpha} + E \left| \frac{\tilde{\epsilon}_0}{\sqrt{n}} \right|^{4\alpha} \right). \end{aligned} \quad (12.12)$$

Since, by the assumptions of the theorem,  $E|\epsilon_0|^p < \infty$  for some  $p \geq \max(6, 4\alpha)$ , we get, by Lemma 12.4, that  $E|\tilde{\epsilon}_0|^{4\alpha} < \infty$ . Since for i.i.d. r.v.'s  $\eta_t$ ,  $t \geq 1$ , and  $p > 2$ ,

$$E \left| \sum_{t=1}^k \eta_t \right|^p \leq K n^{p/2} E |\eta_1|^p \quad (12.13)$$

(see, e.g., Dharmadhikari, Fabian and Jogdeo, 1968, and also de la Peña, Ibragimov and Sharakhmetov, 2003), we also conclude, using Jensen's inequality, that

$$\max_{1 \leq k \leq nN} E \left| \frac{C(1)}{\sqrt{n}} \sum_{t=1}^k \epsilon_t \right|^{4\alpha} \leq \max_{1 \leq k \leq nN} \left( E \left| \frac{C(1)}{\sqrt{n}} \sum_{t=1}^k \epsilon_t \right|^p \right)^{4\alpha/p} \leq L (E|\epsilon_0|^p)^{4\alpha/p}$$

for some constant  $L > 0$ . These estimates evidently imply, together with (12.12), that bound (12.12) indeed holds. ■

## REFERENCES

Akonom, J. and Gouriéroux, C. (1987). A functional limit theorem for fractional processes. Technical Report #8801, CEPREMAP, Paris.

- Anderson, T. W. (1959). On asymptotic distributions of estimates of parameters of stochastic difference equations. *Annals of Mathematical Statistics* **30**, 676–687.
- Avram, F. (1988). Weak convergence of the variations, iterated integrals and Doléans-Dade exponentials of sequences of semimartingales. *Annals of Probability* **16**, 246–250.
- Billingsley, P. (1968). *Convergence of Probability Measures*. Wiley, New York.
- Borodin, A. N. and Ibragimov, I. A. (1995). Limit theorems for functionals of random walks. *Proceedings of the Steklov Institute of Mathematics*, no. 2 (195).
- Burkholder, D. L. (1973). Distribution function inequalities for martingales. *Annals of Probability* **1**, 19–42.
- Chan, N. H. and Wei, C. Z. (1987). Asymptotic inference for nearly nonstationary AR(1) processes, *Annals of Statistics* **15**, 1050–1063.
- Chan, N. H. and Wei, C. Z. (1988). Limiting distributions of least squares estimates of unstable autoregressive processes,” *Annals of Statistics* **16**, 367–401.
- Coffman, E. G., Puhalskii, A. A. and Reiman, M. I. (1998). Polling systems in heavy traffic: A Bessel process limit. *Mathematics of Operations Research* **23**, 257–304.
- Coquet, F. and Mémin, J. (1994). Vitesse de convergence en loi pour des solutions d’équations différentielles stochastiques vers une diffusion. (Rate of convergence in law of the solutions of stochastic differential equations to a diffusion). *Séminaire de Probabilités, XXVIII*, 279–292, Lecture Notes in Mathematics 1583.
- Csörgő, M. and Horváth, L. (1993). *Weighted Approximations in Probability and Statistics*. Wiley, New York.
- Dharmadhikari, S. W., Fabian, V. and Jogdeo, K. (1968). Bounds on the moments of martingales. *Annals of Mathematical Statistics* **39**, 1719–1723.
- Dedecker, J. and Merlevède, F. (2002). Necessary and sufficient conditions for the conditional central limit theorem. *Annals of Probability* **30**, 1044–1081.
- Dedecker, J. and Rio, E. (2000). On the functional central limit theorem for stationary processes. *Annales de l’Institut H. Poincaré. Probabilités et Statistiques* **36**, 1–34.
- de la Peña, V. H., Ibragimov R. and Sharakhmetov, S. (2003). On extremal distributions and sharp  $L_p$ -bounds for sums of multilinear forms. *Annals of Probability* **31**, 630–675.
- de Jong, R. M. (2002). Nonlinear estimators with integrated regressors but without exogeneity. Working paper, Michigan State University.
- Doukhan, P. (2003). Models, inequalities, and limit theorems for stationary sequences. In: *Theory and applications of long-range dependence* (Doukhan, P., Oppenheim, G. and Taqqu, M. S., eds.), 43–100, Birkhäuser, Boston.
- Doukhan, P., Oppenheim, G. and Taqqu, M. S., eds. (2003). *Theory and applications of long-range dependence*. Birkhäuser, Boston.
- Durrett, R. (1996). *Stochastic Calculus. A Practical Introduction*. CRC Press, Boca Raton, FL.

- Dynkin, E. B. and Mandelbaum, A. (1983). Symmetric statistics, Poisson point processes, and multiple Wiener integrals. *Annals of Statistics* **11**, 739–745.
- Giraitis, L. and P. C. B. Phillips (2006). Uniform limit theory for stationary autoregression. *Journal of Time Series Analysis*, 27, 51–60.
- Hall, P. and Heyde, C. C. (1980). *Martingale Limit Theory and Its Application*. Academic Press, New York.
- He, S.-W., Wang, J.-G. and Yan, J.-A. (1992). *Semimartingale Theory and Stochastic Calculus*. CRC Press, Boca Raton.
- Hu, L. and P. C. B. Phillips (2001). Dynamics of the federal funds target rate: a nonstationary discrete choice approach. *Journal of Applied Econometrics* **19**, 851–867.
- Ibragimov, R. and Phillips, P. C. B. (2004). Regression asymptotics using martingale convergence methods, Cowles Foundation Discussion Paper 1473; <http://cowles.econ.yale.edu/P/cd/d14b/d1473.pdf>
- Ikeda, N. and Watanabe, S. (1989). *Stochastic Differential Equations and Diffusion Processes*. North-Holland, Amsterdam.
- Jacod, J. and Shiryaev, A. N. (2003). *Limit Theorems for Stochastic Processes*. 2nd edition. Springer-Verlag, Berlin.
- Jakubowski, A., Mémin J. and Pagès, G. (1989). Convergence en loi des suites d'intégrales stochastiques sur l'espace  $D^1$  de Skorokhod. (Convergence in law of sequences of stochastic integrals on the Skorokhod space  $D^1$ ). *Probability Theory and Related Fields* **81**, 111–137.
- Jeganathan, P. (2003). Second order limits of functionals of sums of linear processes that converge to fractional stable motions. Working paper, Indian Statistical Institute.
- Jeganathan, P. (2004). Convergence of functionals of sums of r.v.s. to local times of fractional stable motions. *Annals of Probability* **32**, 1771–1795.
- Kurtz, T. G. and Protter, P. (1991). Weak limit theorems for stochastic integrals and stochastic differential equations. *Annals of Probability* **19**, 1035–1070.
- Liptser, R. Sh. and Shiryaev, A. N. (1989). *Theory of Martingales*. Kluwer, Dordrecht.
- Mandelbaum, A. and Taqqu, M. S. (1984). Invariance principle for symmetric statistics. *Annals of Statistics* **12**, 483–496.
- Mémin J. (2003). Stability of Doob-Meyer decomposition under extended convergence. *Acta Mathematicae Applicatae Sinica. English Series* 19, 177–190.
- Mémin, J. and Słomiński, L. (1991). Condition UT et stabilité en loi des solutions d'équations différentielles stochastiques. Séminaire de Probabilités, XXV, 162–177. *Lecture Notes in Mathematics* 1485. Springer, Berlin.
- Nze, P. A. and Doukhan, P. (2004). Weak dependence: models and applications to econometrics. *Econometric Theory* **20**, 995–1045.
- Park, J. Y. and Phillips, P. C. B. (1999). Asymptotics for nonlinear transformations of integrated time series. *Econometric Theory* **15**, 269–298.

- Park, J. Y. and Phillips, P. C. B. (2001). Nonlinear regressions with integrated time series. *Econometrica* **69**, 117–161.
- Phillips, P. C. B. (1987a). Time-series regression with a unit root. *Econometrica* **55**, 277–301.
- Phillips, P. C. B. (1987b). Towards a unified asymptotic theory for autoregression. *Biometrika* **74**, 535–547.
- Phillips, P. C. B. (1999). Discrete Fourier transforms of fractional processes. Cowles Foundation Discussion Paper 1243, Yale University; <http://cowles.econ.yale.edu/P/cd/d12a/d1243.pdf>
- Phillips, P. C. B. (2006). Unit root log periodogram regression. *Journal of Econometrics* (forthcoming)
- Phillips, P. C. B. and Magdalinos, T. (2006). Limit theory for moderate deviations from unity. *Journal of Econometrics* (forthcoming)
- Phillips, P. C. B. and Perron, P. (1988). Testing for a unit root in time series regression *Biometrika* **75**, 335–346.
- Phillips, P. C. B. and Ploberger, W. (1996). An asymptotic theory of Bayesian inference for time series. *Econometrica* **64**, 381–413.
- Phillips, P. C. B. and Solo, V. (1992). Asymptotics for linear processes. *Annals of Statistics* **20**, 971–1001.
- Pöetscher, B. M. (2004). Nonlinear functions and convergence to Brownian motion: beyond the continuous mapping theorem. *Econometric Theory* **20**, 1–22.
- Prigent, J.-L. (2003). *Weak convergence of financial markets*. Springer Finance. Springer-Verlag, Berlin.
- Saikkonen, P. and Choi, I. (2004). Cointegrating smooth transition regressions. *Econometric Theory* **20**, 301–340.
- Shorack, G. R. and Wellner, J. A. (1986). *Empirical Processes with Applications to Statistics*. Wiley, New York.
- Stroock, D. W. and Varadhan, S. R. S. (1979). *Multidimensional Diffusion Processes*. Springer-Verlag, New York.
- White, J. S. (1958). The limiting distribution of the serial correlation coefficient in the explosive case. *Annals of Mathematical Statistics* **29**, 1188–1197.